DOCUMENT RESUME

ED 161 312 HE 010 500

AUTHOR	Seibert, Warren F.
TITLE	National Project III, Elevating the Importance of Teaching. Fund Associate's Final Report.
INSTITUTION	Purdue Univ., Lafayette, Ind.
SPONS AGENCY	Fund for the Improvement of Postsecondary Education (DHEW), Washington, D.C.
PUB DATE	Oct 77
GRANT	G007502063
NOTE	80p.; Appendix B not reproducible
EDRS PRICE	MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS	College Instruction; *Computer Oriented Programs; *Course Evaluation; *Educational Assessment; *Evaluation Methods; *Faculty Evaluation; Federal Programs; Higher Education; *Instructional Improvement; Interinstitutional Cooperation; Statistical Analysis; Test Results
IDENTIFIERS	*CAFETERIA System; Purdue University IN

ABSTRACT

Purdue University's participation as a fund associate in National Project III (NP III) for elevating the importance of teaching has its origins in a flexible and diagnostic instructional evaluation system called "CAFETERIA." CAFETERIA services include test development, scoring, and analysis; social surveys on topics of importance in education; assistance in behavioral research and statistical analysis; operation of an instructional grant program; and evaluation of instruction or of educational programs. Local NPIII activities include the continuation of CAFETERIA services to faculty and use of the system by other independent institutions. It is recommended that institutions deciding to implement similar services should operate initially on a limited or pilot basis. Several studies conducted using CAFETERIA data are described. Other NPIII activities include report preparation, participation in colloquia on instruction and evaluation, and aid to other schools in establishing pilot trials of CAFETERIA services. Chances of institutional success in implementing similar evaluation programs are best when there is broad institutional representation in decisions to implement such programs. Evaluation services should be provided directly and in confidence to individual faculty. A first priority need for the future is to achieve a better balance between administrative and diagnostic use of ratings. Appended are a sample of instructor reports; issues of the instructional memo series, "Trends in Learning and Teaching," containing information on the CAFETERIA system; items from faculty evaluation questionnaires; and papers on faculty evaluation and the CAFETERIA model. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED161312

FUND ASSOCIATE'S FINAL REPORT

to

Fund for the Improvement of Postsecondary Education, Department of Health, Education and Welfare, Washington, D.C.

NATIONAL PROJECT III, ELEVATING THE IMPORTANCE OF TEACHING

Grant No. G007502063

Warren F. Seibert
Purdue University
West Lafayette, IN 47907

October, 1977

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

HE 010 500

Contents

	Page
Some Background on the CAFETERIA System	1
The Measurement and Research Center and the University	3
The Status of Teaching at the University	3
Participation in NP III	5
Local NP III Activities	6
Some Studies of CAFETERIA Data	7
Other NP III Activities	11
Some Judgments in Lieu of Conclusions	12
Future Prospects	13
Appendices (A-F)	16 ff.

FUND ASSOCIATE'S FINAL REPORT

NATIONAL PROJECT III, ELEVATING THE IMPORTANCE OF TEACHING

to the

Fund for the Improvement of Postsecondary Education, Department of Health, Education and Welfare, Washington, D.C.

Grant No. G007502063, awarded to Purdue University, West Lafayette IN 47907

SOME BACKGROUND ON THE CAFETERIA SYSTEM

Purdue University's participation as a Fund Associate in National Project III (NP III) has its origins in a flexible and diagnostic instructional evaluation system which is called "CAFETERIA" (see Appendix A). The system was designed in 1972 and, by 1975, when NP III began, it was well established as a service to local faculty. In addition, CAFETERIA services were being provided to faculty at approximately 20 Indiana independent institutions, under terms of a three-year grant from Lilly Endowment; beyond this, the system had been formally adopted by about eight institutions outside of Indiana.

CAFETERIA had itself originated in the desire to overcome several disadvantages of more traditional instructional rating systems or instruments. Few of these instruments yield results that are detailed enough to serve diagnostic purposes. They also lack the flexibility to cope with varied teaching styles and goals. Most, in fact, consist of a single, fixed set of evaluation items and through these they imply a sameness or homogeneity of teaching which CAFETERIA's developers consider unwarranted. The developers also believe that faculty will more readily use evaluation services which they can conform to their own needs and interests. They seem willing to be partners in an evaluation process but are more reluctant and even anxious when confronted with schemes that are unresponsive to their judgments and choices. This is understandable in part because of the faculty's conviction that they are best qualified to state their own instructional purposes, to identify important features of their own instructional styles or strategies, and thus to decide

which questions their own evaluations should address. Similarly, most faculty enjoy the subtleties of language and so, even when they agree substantially on a question or topic, they often disagree on which words convey it best.

It is worth mentioning that there is a tendency for some to regard CAFETERIA as a highly technical and complex system-service. In comparison with other evaluation systems, it may seem that way because its flexibility can give rise to an almost infinite number of unique evaluation instruments. (Other systems typically deal in predetermined sets of perhaps ten to fifty items.) However, as a computer-based system, CAFETERIA manages most detailed processing and record keeping automatically, with the result that these create few burdens. A computer of small to modest size and four specially written computer programs handle each of the following tasks: 1) store a catalog or "library" of available evaluation items; 2) read each instructor's request for specific items to create his/her individualized rating forms; 3) print rating forms in the quantities desired and make a full record of what was done; 4) read students' responses (ratings) from each classroom or similar group, determine which instructor, class, and specific items are being analyzed, and calculate a response median for each item; 5) accumulate specific item medians across groups and construct item norm files from these; 6) prepare a diagnostic report for each instructor (see Appendix B) which lists the contents of each selected item, its response distribution, median, and percentile equivalent; 7) prepare a mailing label which addresses each instructor report; and 8) merge accumulated data from two sources to create enlarged or updated norm files.

The computer also routinely includes certain contents on each rating form. Besides identifying the instructor and course, each form includes several student-course demographic items (e.g., year in school, sex, expected final grade) and several global "core" evaluation items. Locally, a set of five core items appears on every form. These elicit students' judgments of the instructor's ability to motivate learning, ability to explain clearly, quality of course assignments, overall quality of the

course, and overall quality of the instructor. Beyond the processes which are computer controlled, there is of course some manual, clerical processing, but this is relatively limited and all such tasks are so defined that they qualify as simple routines.

THE MEASUREMENT AND RESEARCH CENTER AND THE UNIVERSITY

The CAFETERIA system was developed by the staff of Purdue's Measurement and Research Center, a unit which serves faculty and others in the University community. Its principal responsibilities can be classified as forms of applied psychology: test development, test scoring and test analysis, social surveys on topics of importance in education, assistance in behavioral research and statistical analysis, operation of an instructional grant program, and evaluation of instruction or of educational programs. In 1975 and for some years prior, MRC was responsible to the University's Vice President for Student Services; however, following a reorganization in 1976, the Center became responsible to the Associate Provost and Dean of Academic Services. This administrative relocation now places the Center among most other academic or faculty service units, including Telecommunications (ITV), the University's public radio station, Libraries-Audio Visual Center, Computing Center, a small Office of Instructional Materials Development, and Conferences-Continuing Education. The larger setting for the Center and for the present project is that of a public, land grant university which enrolls 30,000 students on its main campus. (Included are 5,500 graduate students.) The faculty numbers 1,700 in the professorial ranks, about 70 percent of whom are tenured. An additional 15,000 students attend four other University campuses.

THE STATUS OF TEACHING AT THE UNIVERSITY

It seems particularly helpful for present purposes to consider also the status (the importance) of teaching at the University. While there are no precise answers to such questions, there are indications that seem instructive. Not surprisingly, there are important differences among

the nine schools and numerous departments. In general, professional or vocational schools and programs emphasize teaching responsibilities more than do schools or departments which represent traditional disciplines. In addition, there are signs of a broad move which is increasing the emphasis on teaching. Nevertheless, this appears not to affect greatly the traditional emphasis on research, scholarly work, and creative productivity. Thus, it appears that faculty are increasingly expected to perform well both as teachers and as scholars or creators. In some respects, such versatility has always been expected of the faculty, but a decade and more ago it was easier to compensate with research for some inadequacy in teaching; compensation in the opposite direction is not unprecedented but still occurs less frequently. To explain these apparent changes, we would invoke the same points that others frequently make: concern over education and the quality of teaching is an insistent issue, locally and nationally, which has come to require response; enrollments at the University remain high and are in fact still increasing slightly, so there still are many students to be taught; inflation is outpacing the availability of funds, so new hiring to cover expanded duties is rarely possible; and the academic job market offers faculty few options except to make the best of existing circumstances.

The increasing emphasis on teaching became most evident to the writer and his colleagues soon after CAFETERIA services were introduced. In the Fall of 1972, evaluations from 400 classrooms were processed, then in succeeding semesters the numbers were 580, 710, 760, 1300, and 1420. Since then, local uses of the service have fluctuated in the 1250 to 1750 range, high in the fall, lower in the spring semester, and these levels now correspond to 25-35 percent of the course sections taught during each semester. No one had expected such usage to develop so quickly, if at all, and the statistics stand as reasonably good evidence of faculty interest in the quality of their teaching. "What kind of interest?" is another matter and more difficult to answer.

Until perhaps 1976, it was plausible that most who used the service were interested primarily in diagnostic feedback and, in the purest cases, in instructional "fine tuning". They were, in other words, using CAFETERIA just as its developers had hoped and intended. Now, such views can be challenged. And, while there are many known examples of diagnostic use, there also are clear signs that CAFETERIA results are regularly presented and considered in reviews for faculty promotion and tenure. Such use will be discussed later, together with several issues that it raises. For now it will suffice to say that CAFETERIA services are widely used, the use seems both a symptom and a cause of increased attention to effective teaching, and it raises issues which deserve further examination in a later section of this report.

PARTICIPATION IN NP III

When NP III began, it was difficult to know what to expect of it or, in turn, exactly what was expected of a Fund Associate. The theme of NP III, Elevating the Importance of Teaching, was one which coincided with local CAFETERIA activities and so the expectation was that the University would benefit by learning of related activities elsewhere, that it would be able to refine its instructional evaluation and improvement efforts, and that generalizable features of local efforts could become known to a larger audience and be scrutinized by this audience. Thus, we expected opportunities to benefit from specifics of other instructional programs, to consider important issues which such programs raise, and to communicate gleanings from local experience to others in higher education. We expected, too, that it would be interesting to participate in an experiment of sorts, since the "self-destructing" consortium which NP III represents had no known precedent; could such a group, of unpredictable size and composition, be created, develop an agenda, do its business, and dissolve, all in the space of about two years? (The brief answer, we believe, is yes.)

This Fund Associate's NP III activities divide conveniently, if not precisely, into two groups: 1) those that were primarily local in

implementation or immediate impact and 2) those executed off-campus or related more to colleagues elsewhere, including other Fund Associates. In considering local activities, crisp distinctions cannot be made between those that were formally within the project and those that were simply on-going but relevant to project purposes. For this report, the best criterion is relevance and it will therefore be employed.

Local NP III Activities

Among the local activities, there was of course the continuation of CAFETERIA services to faculty and, during four semesters and two summer sessions, more than 6000 evaluations ("classrooms") were processed. This translates into approximately 150,000 student rating forms which were prepared and analyzed. For sheer volume and accumulation of experience, this probably equals or exceeds that of any institution. Additional and valuable experience was gained from the continuation through 1976 of the Lilly-funded project and from the growing number of other institutions which were adopting CAFETERIA. As mentioned above, institutions participating in the Lilly project numbered about 20 and all are independent colleges or universities within Indiana; the majority are liberal arts institutions, although a few are more specialized and vocational in their curricula. During the three semesters prior to the conclusion of that project (a time period which overlaps the activities of NP III), evaluations from 3700 courses were processed; this corresponds to about 90,000 student rating forms.

Faculty at the independent institutions demonstrated much the same readiness to use the service that had been noticed locally; it also appeared that, in general, the service fared best at institutions which organized, deliberated, and prepared for its use. Or, on the other hand, institutions which acted too rapidly, perhaps by edict rather than through consultation and involvement of faculty, were more likely to encounter later problems. Among other adopting institutions, which now total about 45, the experience seems similar. So, while it is no great discovery, our consistent recommendation to institutions contemplating such a ser-

vice is that local decisions be reached as a cooperative undertaking involving representatives of all interested groups. More specifically and concretely, institutions which decide to inaugurate such services are encouraged to operate initially on a limited or pilot basis and thus provide time for faculty to acquire familiarity gradually and time also for local staff to refine their operating procedures and their roles. We also urge institutions to preserve as much flexibility of use as they can, since experience suggests that faculty interest can decline as elective options in evaluation are restricted.

Experience and suggestions like those mentioned above have been regularly shared with local groups and individuals and with approximately 200 others who have made specific enquiries by mail during the period of NP III. Similar information is presented in articles disseminated to all faculty (see Trends in Learning and Teaching, Appendix C), in a special meeting of MRC staff with the University's academic deans, and in invited presentations before most school and several departmental faculties.

Some Studies of CAFETERIA Data

Data available from large scale local uses of CAFETERIA provide many opportunities to study characteristics or qualities of students' ratings, so although the first objective of the staff has been service rather than research, several studies have been conducted and a few of them deserve brief mention here.

Questions of reliability have been examined in several ways. The earliest study employed Cronbach's alpha statistic, yielding an interval consistency estimate. It found an average reliability for results from CAFETERIA forms to be .88. Later, a form of test-retest estimate was obtained from several situations in which instructors used a common rating form (i.e., common items) in two sections of the same or a similar course. The profiles formed by item median ratings appear usually to correlate about .80. Thus, the strengths and weaknesses identified in one section coincide well with those in the other.

In addition, it has been of interest to examine matrices of item intercorrelations, since the higher correlations in these matrices can be regarded as lower bound estimates of individual item reliability. (In other words, excepting the occasional chance occurrence, a measure will not correlate more highly with some other measure, i.e., with another item, than with itself.) One such matrix was calculated in September 1975 with data from 1032 classrooms and it reveals a fair proportion of inter-item correlations in the range of .70 to .85. Also, looking only at the five core items*, their intercorrelations were:

Core Item	2	3	4	5
1	.84	.77	.88	.87
2		.66	.70	.86
3			.89	.70
4				.79

The consistency of these same ten correlations was examined with comparable data from a preceding semester, representing about 1150 cases. All but one of these coefficients was within .03 of the values given above (five were identical or within .01) and the "deviant" case differed by .05, a decline from .88 to .83.

It should be mentioned that the correlations just cited are not to be regarded as typical of all lower bound estimates, since CAFETERIA's core items have the advantage of reasonably normal distributions and broad score ranges. A number of the other items suffer from at least some skewness and range restriction, both of which operate to reduce the size of obtainable correlations. Nevertheless, the indications from several sources are that the rating results provided to faculty satisfy generally accepted standards of reliability.

Parenthetically, the generally high and consistent correlations among the five core items are a mixed blessing. As lower bound esti-

---

*The contents of the five items, in order, are "My instructor motivates me to do my best work", "My instructor explains difficult material clearly", "Course assignments are interesting and stimulating", "Overall, this course is among the best I have ever taken", and "Overall, this instructor is among the best teachers I have known".

mates of the items' reliability, they add to the conviction that these items yield stable results. But they have a further meaning which is not highly desirable. Since the correlations tend to be high, the items' results tend to be repetitive, rather than yielding more independent indices of reasonably unique instructional features. By squaring each of the correlations, one learns that the shared variance between pairs of items ranges between 44 and 80 percent. Or, stated as an average, 60-65 percent of the variance of two items is common and this translates into a correlation of almost .80. Then, if corrected for the items' unreliability, the typical correlation would rise still higher, probably into the range of .85-.90.

Two post-doctoral investigators* were invited to spend about six months working with the MRC staff on a variety of exploratory analyses and studies of CAFETERIA data. They assisted with portions of the work described above and conducted exploratory analyses which are briefly mentioned at the close of this section, but their principal effort consisted of a structural analysis of CAFETERIA ratings, using the methods of non-metric multidimensional scaling (NMDS) and heirarchical cluster analysis.

Working with CAFETERIA data from 1841 classrooms, a group of 20 departments was first identified in which faculty had made greatest use of CAFETERIA services. Data referring to those departments were then used to identify 83 CAFETERIA items which these faculty had most frequently selected; these "modal" items accounted for almost 75 percent of all item usage in the original data pool. Then, since the NMDS computer program could accommodate no more than 61 items at a time and since it is desirable to minimize the number of empty cells in the correlation matrix, some further items were deleted. The resulting 61 x 61 matrix had 18 empty cells, of a total of 1830 and each correlation was based on at least 50 observations.

---

*They were Dr. Raanan Lipshitz, now of Haifa, Israel and Dr. William M. Goodwin III, now of the University of Miami, Coral Gables, Florida.

The NMDS analysis of these correlations resulted in a six-dimensional solution which was then subjected to heirarchical cluster analysis, yielding ten identifiable, first-order item clusters. These were named as follows: instructor fairness, instructor helpfulness, instructor preparedness, global instructor effectiveness, global course challenge, course applicability, course format, clarity of course goals, instructional pace, and instructional climate. (For the contents of items that constitute these clusters, see Appendix D.) In a higher order cluster, preparedness and applicability combine with global course and global instructor clusters and form a complex which appears to represent students' satisfaction with the instructional experience.

Also, instructional pace or reasonableness combines with climate to form an intermediate level cluster, as do fairness and instructor helpfulness. Reasonableness is briefly discussed below and climate represents a clustering of items which deal with freedom to express oneself in class, to raise questions, sufficiency of time for questions and discussions, and encouraging students' contributions to class learning, among others. The other intermediate cluster includes fairness, which is predominantly a matter of fair examining and grading practices, plus two more general items that refer to general instructor impartiality and openness of learning opportunities, plus the instructor helpfulness or, perhaps, empathy cluster. The four items of the latter set contain references to instructor sensitivity to students' problems, providing help where needed, allowing challenges of an instructor's ideas, and returning papers quickly enough to benefit the student.

These clustering results are not considered to be novel, since a number of earlier studies have reported similar factors or clusters, but one cluster strikes the writer as most interesting and as an especially worthy challenge to every serious teacher. It bears the innocuous label, preparedness, but what the contents of the clustered items reveal is a portrait of a teacher who maintains an artist's commitment to his/her work. It includes being well prepared for class and displaying a clear understanding of course topics, but also care and precision in answering

questions, ability to simplify difficult material, maintaining rapport and holding class attention, and speaking clearly. To do well on these things is, we submit, to display the highest sense of responsibility toward one's students and one's discipline. On the other hand, our candidate for the least noble cluster is the one labelled instructional pace, but formerly called "reasonableness". Its items include references to the student being able to keep up with the work load, to complexity and length of assignments as reasonable, to the amount of covered material as reasonable, and to length and difficulty of assigned readings as reasonable. One wonders whether all that reasonableness is not, in the light of education's higher goals, unreasonable. Nevertheless, the results of these analyses are further evidence that students' ratings are not simply based on instructor popularity, as critics sometimes contend. Instead, students can and do make a variety of discriminating judgments about aspects of their instructional experience, and they make these judgments reliably.

There have been a number of other studies and exploratory analyses, but these either were inconclusive or they have little importance, except locally. They examined such questions as the influence of student or course characteristics (e.g., size of course enrollment, course level, student group composition, school in which a course was offered) upon the level of the core item ratings. The principal result of this work is that discriminated item norms are now under development (see p. 13).

Other NP III Activities

Several NP III activities have involved each Fund Associate in tasks which are primarily intended to serve colleagues at other institutions. In our own case, we participated in each of the Fund Associate meetings and shared responsibility for editing or authoring major sections of most CRITERIA reports; our largest role was in the preparation of reports numbered II and VI. In addition, CAFETERIA staff participated in colloquia or symposia on instruction and evaluation at about two dozen colleges or universities, two FIPSE project directors' conferences, and two national conventions (See Appendices E and F). We also helped to esta-

blish pilot trials of CAFETERIA services at three Fund Associate campuses, Ohio Wesleyan, Illinois, and Bucknell. Each had expressed interest in trying the service on a small scale and, at this writing, one has adopted it, one is creating a local variant, and the third appears to have made no decision. One other Fund Associate, Earlham College, was already making some use of CAFETERIA services as a participant in the Lilly Endowment project.

SOME JUDGMENTS IN LIEU OF CONCLUSIONS

Experience during the two years of NP III and the studies based on CAFETERIA data provide a basis for several judgments which others may find useful. For the most part these judgments are not beyond dispute. They are, rather, opinions developed from extensive and varied experience. First, as mentioned above, we strongly suspect that the chances of institutional success in implementing similar evaluation programs are best when there is broad institutional representation in decisions to implement such programs. Similarly, to the extent that flexibilities or individual options are preserved, constructive faculty participation seems most likely to follow. This judgment can be defended best by describing an extreme case: Consider the example of an institution at which a peremptory decision is made to require that all faculty use a given evaluation instrument and under highly specified conditions. Few institutions could fully implement such an edict, even if they choose to try, and fewer still would find the benefits to be greater than the costs. As a matter of principle and perhaps of temperament, faculty resist edicts. Such resistance seems especially great when decisions impinge upon or seem to impugn the faculty's teaching performance.

Secondly, the rationale of CAFETERIA (see p. 1 and Appendix A) is one which seems as attractive now as it did at the outset. The broadest of the guiding principles is that evaluation services be provided directly and in confidence to individual faculty. In other institutional climates a different conception may prove workable, but in spite of several local challenges to the principle, it continues to seem the best course at this University. Through it, faculty are reassured that the

CAFETERIA (MRC) staff is interested only in helping with the faculty's efforts to improve instruction and in encouraging the expansion of such efforts.

A central reason for maintaining the principle of confidential, individual service and for emphasizing diagnosis is that there appear to be ample pressures and activities of an opposing kind. It is clear, for example, that many faculty now use their CAFETERIA results to help demonstrate adequacy or excellence in teaching and that, in many quarters, certain pressures are exerted to cause faculty to "share" their results, typically with department heads. These uses are largely incompatible with diagnostic, instructional improvement uses, so the preservation of diagnosis as a credible and principal purpose is in some danger.

Thus, experience suggests that an analog of Gresham's law is at work and one can doubt that any institution will be immune to it. Diagnosis and improvement are noble activities which may also bring tangible rewards to faculty, but administrative decisions on promotion and salary deal directly in these rewards. Also, the problem of competing uses would be less serious if more faculty and administrators discriminated carefully between them. But since, on the surface, the results look much the same, their uses become easily confused. The dilemma in this is that students' ratings have definite value in deciding questions of teacher effectiveness. Thus, the CAFETERIA staff cannot in good conscience recommend against their administrative use and, so far, it has found no adequate means to balance this use with what it regards as the more important, more constructive, and longer term values of diagnosis.

FUTURE PROSPECTS

To conclude, a few words on future prospects seem in order. As already implied, a first priority need is to develop by some means a better balance between administrative and diagnostic use of ratings. Also, work is proceeding on the development of discriminated item norms. Currently and in the past, the University has employed one undifferentiated set of norms for each CAFETERIA item. The desire has been to

develop several sets, each based on similar or fairly homogeneous groups of courses or instructors. If not for the task of "cleaning" data first, this project could be accomplished in a matter of days or weeks, but under the prevailing circumstances, it is more realistic to predict that several months are required.

It is perhaps of interest that plans have been prepared for developing a system based on CAFETERIA but directed toward needs in continuing and adult education (CAE). Colleagues have convinced the staff that CAFETERIA's main features are well suited to needs in CAE and several exploratory uses of the system have given further confirmation. Also, as we interpret the literature in the field, CAE is a growing and increasingly important sector within education and its need for improved evaluation practices is equally clear. It cannot be predicted when this new developmental effort can be fully launched or completed, but preliminary work is underway.

Finally, the outlines have been developed for a continuing study, the purpose of which would be to monitor trends that bear on the importance of teaching. Whether complete plans and suitable funding for the study can be developed remain to be seen; however, if they can, then the lens model of the late Egon Brunswik would be employed in a series of (probably) annual studies, each of which yields estimates of the importance accorded to teaching by decision makers in postsecondary education. Together, studies in the series should reveal trends of increasing or lessening importance, plus much further information on the status and trends of other factors which decision-makers employ. To some, such studies are regarded as policy capturing, since they analyze information available to decision-makers (e.g., members of departmental promotions committees) and subsequent decisions of these groups, then reach conclusions concerning the factors which influence decisions. In this case, how important is teaching? Current and expected future conditions in higher education include a number of new influences or older influences which are assuming new importance. No one can predict how these will combine to alter the future consideration of teaching in postsecondary

institutions. We believe, therefore, that it is increasingly important to monitor those actions which, when taken, serve either to elevate or diminish the importance of teaching.

CAFETERIA Course-and-Instructor Evaluation System

Beginning in early 1972, several staff members of Purdue's Measurement and Research Center undertook to develop a new and flexible system of course-and-instructor evaluation. The product of their efforts is now known as "the CAFETERIA system". The most general purpose of these efforts was to overcome weaknesses that were thought to be common among many of the more traditional instructional rating schemes. Particular emphasis was given to the following:

1) A perceived need for detailed and diagnostic evaluation results, rather than more general or global evaluations, so that faculty could use results directly in the improvement of teaching strategies.

2) A need to involve faculty in the shaping or design of their own evaluation instruments in order to gain wide faculty acceptance, to gain increased sensitivity to important course and teacher characteristics, and to achieve increased salience of the results.

3) A need for content flexibility in evaluations, not only to meet the unique requirements of essentially traditional courses and teachers but also to satisfy the requirements of unusual or non-traditional teaching.

4) A great practical need for efficient operation, so that large volumes of tailored service could be achieved, yet manageably and inexpensively. This clearly required the use of a computer.

5) A need to minimize the threat that evaluation may pose for teachers, primarily by stressing the diagnostic, instructional improvement purposes of the service and by guarding the confidentiality of results.

6) A need to focus increased attention on the importance of effective college and university teaching, by whatever ethical means are available.

7) A need to recognize and develop some response to demands for teacher evaluations that could be of some use in personnel review and administrative decision-making.

The CAFETERIA system that was evolved, and is still evolving, meets most if not all of the original design objectives. In typical operations, faculty are at first invited to participate in the service and they receive a catalog (or "menu") of instructional rating items and an item selection form. Faculty may select as many as forty items from

the catalog by marking their choices on the special form. Then, when this form is processed, the computer line printer prepares rating questionnaires, in desired quantities, and these are forwarded to the faculty member for administration in class. Faculty return the completed questionnaires for processing and they then receive two copies of the confidential, diagnostic report. (Samples of the item selection form, instructional evaluation questionnaire, and instructor report are attached.)

Appendix B

SAMPLE INSTRUCTOR REPORT

05/27/77

STUDENT DEMOGRAPHICS -

CLASS		SCHOOL				EXPECTED GRADE		COURSE REQUIRED	
FRESHMAN	0	AGR	1	SCI	4	A/PASS	25	YES	41
SOPHOMORE	1	ENGR	27	TECH	1	B	16	NO	3
JUNIOR	40	HOME EC	0	VET SCI	0	C	2		
SENIOR	3	HSSE	2	SEX		D	0		
GRADUATE	0	IM	8	FEMALE	3	F/FAIL	0		
OTHER	0	PHARM	0	MALE	38				

CATALOG NUMBER			SA (5)	A (4)	U (3)	D (2)	SD (1)	MEDIAN	PER CENTILE
2	1.	MY INSTRUCTOR DISPLAYS A CLEAR UNDERSTANDING OF COURSE TOPICS.	27	16	0	0	0	4.7	58
6	2.	MY INSTRUCTOR HAS AN EFFECTIVE STYLE OF PRESENTATION.	16	25	2	0	0	4.3	69
9	3.	MY INSTRUCTOR SPEAKS AUDIBLY AND CLEARLY.	28	15	0	0	0	4.7	76
14	4.	MY INSTRUCTOR HOLDS THE ATTENTION OF THE CLASS.	19	22	2	0	0	4.4	78
16	5.	MY INSTRUCTOR STIMULATES INTEREST IN THE COURSE.	10	24	8	1	0	4.0	53
17	6.	MY INSTRUCTOR DISPLAYS ENTHUSIASM WHEN TEACHING.	10	26	4	1	0	4.1	36
18	7.	THIS COURSE SUPPLIES ME WITH AN EFFECTIVE RANGE OF CHALLENGES.	6	19	11	6	1	3.7	15
25	8.	MY INSTRUCTOR HAS STIMULATED MY THINKING.	11	21	7	4	0	4.0	54
26	9.	MY INSTRUCTOR HAS PROVIDED MANY CHALLENGING NEW VIEWPOINTS.	12	20	9	2	0	4.0	58
27	10.	MY INSTRUCTOR TEACHES ONE TO VALUE THE VIEWPOINT OF OTHERS.	8	24	8	3	0	3.9	42
35	11.	MY INSTRUCTOR ENCOURAGES STUDENT CREATIVITY.	7	20	13	3	0	3.8	17
38	12.	MY INSTRUCTOR EMPHASIZES CONCEPTUAL UNDERSTANDING OF MATERIAL.	12	23	7	1	0	4.1	49
41	13.	MY INSTRUCTOR MAKES GOOD USE OF EXAMPLES AND ILLUSTRATIONS.	10	25	6	1	1	4.0	34
44	14.	MY INSTRUCTOR IS ACTIVELY HELPFUL WHEN STUDENTS HAVE PROBLEMS.	21	17	5	0	0	4.5	60
50	15.	MY INSTRUCTOR IS CAREFUL AND PRECISE WHEN ANSWERING QUESTIONS.	22	20	0	1	0	4.5	89
51	16.	MY INSTRUCTOR IS READILY AVAILABLE FOR CONSULTATION.	26	16	1	0	0	4.7	87
52	17.	MY INSTRUCTOR REGULARLY CHECKS AND REWARDS PROGRESS IN LEARNING.	3	17	22	0	1	3.4	19
53	18.	MY INSTRUCTOR SUGGESTS SPECIFIC WAYS I CAN IMPROVE.	9	21	12	1	0	3.9	30
61	19.	MY INSTRUCTOR TAILORS THIS COURSE TO HELP MANY KINDS OF STUDENTS.	4	17	16	5	1	3.5	23
64	20.	I WAS ABLE TO KEEP UP WITH THE WORK LOAD IN THIS COURSE.	10	26	4	3	0	4.1	57
68	21.	I AM FREE TO EXPRESS AND EXPLAIN MY OWN VIEWS IN CLASS.	34	9	0	0	0	4.9	96
69	22.	WHEN I HAVE A QUESTION OR COMMENT I KNOW IT WILL BE RESPECTED.	27	15	0	1	0	4.7	81
76	23.	MY INSTRUCTOR RELATES TO ME AS AN INDIVIDUAL.	25	17	1	0	0	4.6	77
78	24.	MY INSTRUCTOR READILY MAINTAINS RAPPORT WITH THIS CLASS.	26	16	1	0	0	4.7	82
84	25.	I UNDERSTAND WHAT IS EXPECTED OF ME IN THIS COURSE.	21	20	2	0	0	4.5	60
90	26.	THIS COURSE MATERIAL IS PERTINENT TO MY PROFESSIONAL TRAINING.	19	11	6	4	1	4.3	54
94	27.	MY TECHNICAL SKILLS WERE IMPROVED AS A RESULT OF THIS COURSE.	2	13	19	5	4	3.2	4
106	28.	THERE IS SUFFICIENT TIME IN CLASS FOR QUESTIONS AND DISCUSSIONS.	23	17	2	0	1	4.6	86
112	29.	THIS COURSE PROVIDES AN OPPORTUNITY TO LEARN FROM OTHER STUDENTS.	16	24	2	1	0	4.3	55
113	30.	EXAMS ACCURATELY ASSESS WHAT I HAVE LEARNED IN THIS COURSE.	14	19	7	2	1	4.1	89
117	31.	EXAMS STRESS IMPORTANT POINTS OF THE LECTURES/TEXT.	17	23	2	0	1	4.3	82
122	32.	EXAMS ARE COORDINATED WITH MAJOR COURSE OBJECTIVES.	13	20	8	1	1	4.1	59
125	33.	GRADES ARE ASSIGNED FAIRLY AND IMPARTIALLY.	16	22	5	0	0	4.3	76
126	34.	THE GRADING SYSTEM WAS CLEARLY EXPLAINED.	16	22	2	1	0	4.3	76
136	35.	DIRECTIONS FOR COURSE ASSIGNMENTS ARE CLEAR AND SPECIFIC.	18	21	2	2	0	4.3	79
137	36.	THE NUMBER OF COURSE ASSIGNMENTS IS REASONABLE.	12	26	3	1	1	4.1	64
141	37.	STUDENT PRESENTATIONS SIGNIFICANTLY CONTRIBUTE TO THIS COURSE.	20	15	3	3	2	4.4	56
188	38.	I WOULD ENJOY TAKING ANOTHER COURSE FROM THIS INSTRUCTOR.	23	16	3	1	0	4.6	72
192	39.	THESE ITEMS LET ME APPRAISE THIS COURSE FULLY AND FAIRLY.	12	21	6	4	0	4.0	87
196	40.	MY INSTRUCTOR IDENTIFIES MAJOR OR IMPORTANT POINTS IN THE COURSE.	18	23	1	1	0	4.3	70
CORE 1	41.	MY INSTRUCTOR MOTIVATES ME TO DO MY BEST WORK.	10	20	10	3	0	3.9	60
CORE 2	42.	MY INSTRUCTOR EXPLAINS DIFFICULT MATERIAL CLEARLY.	16	24	3	0	0	4.3	82
CORE 3	43.	COURSE ASSIGNMENTS ARE INTERESTING AND STIMULATING.	2	16	6	12	7	2.9	16
CORE 4	44.	OVERALL, THIS COURSE IS AMONG THE BEST I HAVE EVER TAKEN.	7	12	14	7	3	3.3	41
CORE 5	45.	OVERALL, THIS INSTRUCTOR IS AMONG THE BEST TEACHERS I HAVE KNOWN.	18	15	9	1	0	4.3	72

REPORT BASED ON 43 STUDENTS. 05/27/77 COPYRIGHT 1975 PRF.

Appendix C

Issues of the instructional memo series, Trends in Learning and Teaching, each containing information on the CAFETERIA instructional evaluation system.

Volume 1

Number 1

October 1975

Trends in Learning & Teaching

A memo series for Purdue's teaching faculty, issued twice each semester.

THIS NEW SERIES: TRENDS IN LEARNING AND TEACHING

For some time the feeling has grown that a memo series such as this one would fill a need expressed by Purdue's teaching faculty. Such a series would address itself to a variety of topics in the broad range of instruction and learning and would be both a service in its own right and a promoter of the campus's other established services to instruction. The need has been most apparent when talking with faculty about newer developments in instruction and about how faculty might take advantage of these developments.

In launching the series, we have ideas about where it may go and what it may report, but editorial policy remains flexible. By exercising this flexibility and by trying various ideas and memo contents, so long as they pertain to instruction and learning, we expect clearer policy will evolve. Nevertheless, it would be no surprise to see two or three things develop. First, the series may come to resemble series already well established at the University of Michigan, where instructional development and memo series responsibility are focussed in the Center for Research on Learning and Teaching and at McGill in Canada, where The Center for Learning and Development has similar responsibilities. Both of their series are issued a few times each year and each issue includes brief articles and bibliographies on various aspects of the instructional process. They at times deal directly with "innovative teaching", but even when they don't, innovation can often be seen beneath the surface.

A related possibility is that the issues in this series will come to include material of four fairly broad kinds: 1) current or topical information, such as instructional grant program deadlines, awards, and reports of upcoming speakers or other events; 2) broader and more conceptual articles on the instructional process, e.g., teaching for transfer, instructional diagnosis, or the individualization of instruction; 3) reports on exemplary programs of instruction and, especially, successful local efforts in teaching and in solving troublesome instructional problems; and 4) practical or how-to-do-it pieces on such things as using services of Purdue's Film Library, securing TV materials prepared elsewhere, or evaluating new courses.

Your suggestions for the series are welcome and will be used to make subsequent issues more responsive to the most prevalent needs and interests of the Purdue teaching faculty.

THE FILE FOLDER

With this memo you have received a file folder that we hope will be doubly useful, first as a convenient way to collect this and the later issues of this memo series, but also as a directory to Purdue's instructional services and its instructional support units. The directory bears some resemblance to the "Yellow Pages" of a telephone book and it is meant to supersede the old (1971) Educational Equipment and Aids Directory: A Handbook for the Faculty of Purdue University. Besides being out of date, the old Directory did not seem as handy as a handbook ought to be. In any event, we hope you will find value in both the directory and the collecting-filing functions of the folder.

PURDUE'S OCTOBER CONFERENCE ON TEACHING

In late October (27-31), Purdue will hold its third open conference on teaching, sponsored by the Faculty Committee on Educational Equipment and Aids and the Office of Instructional Services, with support from the President's Office. As in past years, the conference will highlight new approaches to instruction and will be addressed principally to the Purdue teaching faculty. Details are being developed now, but the main conference events have been confirmed. Dr. W. James Popham of UCLA will keynote the conference with a 4:00 p.m. presentation in Fowler Hall, on Monday, the twenty-seventh. Dr. Popham is nationally known for his work and writings on instruction, instructional objectives, and evaluation of instruction. Purdue's own Professor Samuel Postlethwait will make a 7:30 p.m. presentation in Fowler Hall on Wednesday, October twenty-ninth. As many know, Dr. Postlethwait is nationally and internationally recognized for his pioneering work on a form of individualized teaching called Audio-Tutorial Instruction. His presentation will be concerned with recent developments in individualized instruction.

Throughout the conference period, participants will have opportunities for site visits-discussions with local colleagues engaged in a variety of innovative instructional improvement efforts. These will include computer-assisted instruction, "Keller Plan" (or PSI), instructional simulation, the CAFETERIA evaluation system, and many other topics.

Please make a note of the conference dates. Through campus mail you will soon receive the conference announcement brochure. In the meantime if there are any questions, please call the Office of Instructional Services, 'phone 2207.

CAFETERIA evaluation data is read from sheets onto tape (above) for complete processing.

THE CAFETERIA INSTRUCTIONAL EVALUATION SERVICE

Purdue's CAFETERIA course-and-instructor evaluation system will undoubtedly be a topic in several later issues of this memo, so we will try not to dwell on it here. But we should at least indicate the purposes and main features of CAFETERIA and should make certain too that all faculty know of the service that CAFETERIA represents. Two things deserve special emphasis: 1) CAFETERIA's results are diagnostic and they are meant principally to serve the individual teacher interested in identifying instructional strengths and weaknesses; 2) CAFETERIA results (instructor evaluation reports) are treated always as confidential information. When a faculty member desires that someone else - a colleague, promotion committee, or whoever - see his/her confidential report, as quite a few seem to do, they make their own provisions for this. Recognizing this interest of faculty, two copies of each confidential report are routinely prepared and provided to participants.

The main features of CAFETERIA and the arguments for them have been stable for about three years now. Even the system's details have remained unchanged for about a year, while developmental work has continued behind the scenes. Some changes will probably be introduced within the next year, as significant improvements are more clearly identified, but it has seemed important to continue a stable service that faculty could try and could become familiar with, and that essentially is part of CAFETERIA policy.

In designing CAFETERIA, the goals of a diagnostic, confidential service were primary. In addition, the following things were sought: methods to encourage wide faculty adoption, especially by involving each participant in the process of questionnaire (or rating instrument) design; questionnaire contents that could encompass unusual as well as normal or customary varieties of courses and teaching; simplicity and efficiency of operation, to minimize the instructor's burden and (selfishly) the burden on the service unit that processes many evaluation results; and finally, some utility in personnel review situations where faculty might voluntarily submit evaluation reports as evidence of teaching effectiveness. Even though CAFETERIA's design achieves each of these things quite well, well enough in fact to have been adopted by McGill, Michigan, California-Davis, and nearly 40 other institutions in Indiana and elsewhere, there are few illusions about the extent of the validity achieved or achievable and, as a respected colleague puts it, "Teaching is still a very human and individual endeavor and its meaning is not easily captured by statistics."

All Purdue faculty and teaching assistants have an open invitation to use CAFETERIA's services. The necessary materials are available at all times and can be obtained from the Measurement and Research Center (ENAD 402) or by dialing 2207 and asking for Ms. Marjorie Cree. Requests for service can be filled within five working days and, once the results are returned to MRC for analysis, about the same processing time is needed to prepare the confidential report copies. During non-peak periods, these turnaround times may be greatly reduced. One further item should be mentioned: an increasing number of faculty desire some help in interpreting their diagnostic reports and the MRC professional staff has gained some experience in providing these interpretations. Interested faculty can make appointments to see either Dr. J.O. Derry or Dr. W.F. Seibert by telephoning 2207.

Confidential CAFETERIA evaluation report is mailed to a participating faculty member.

HELP WITH CLASSROOM TESTING AND EXAMINING

Purdue's Measurement and Research Center has traditionally provided a large volume of service in the scoring and analysis of the faculty's classroom tests. This service continues and is presently in the process of being considerably improved; a later announcement will fully describe the improved service. Essentially though the present service consists of the capability to score objective tests (usually, multiple choice and/or true-false) and to provide the related test and item statistics. To use the service, student responses must be recorded on a special computer response card; this card can accommodate a maximum of five alternatives per test item and a maximum (using two cards) of 200 items. The instructor then receives not only scores for all students and part- or sub-test scores, if desired, but extensive statistics on the functioning of all test items; their level of difficulty, their correlation with total score, student use of each alternative response, and an internal consistency reliability estimate for the test.

The improved service will soon allow the use of a more versatile test response sheet, in place of the response card. The sheet will accommodate a maximum of 80 test items, each with a maximum of ten alternatives per item (which facilitates use of so-called "matching" items) or can be extended to a 160 item maximum when there are no more than five alternatives per item. The projected date for inauguration of this service is mid-November.

On a related matter, faculty frequently seek guidance in developing better classroom tests and, as one response to this, Measurement and Research Center can now provide some materials that have been specially prepared for that purpose. This classroom testing "kit" includes four practical booklets: on making the classroom test, selecting an achievement test, on multiple choice question development, and short-cut test statistics. The kit was authored by staff of the Educational Testing Service, Princeton, and limited quantities are available from Mr. John Van Horn, MRC, 'phone 2207.

* *

From the pen of the late E. G. Boring, dean of American Psychology and a Professor at Harvard:

"...persuasion and shaping the behavior of others is the chief social activity of the gregarious human race, and most of it is hidden! The ZEITGEIST that prevents men from having original thoughts before the culture is ready to receive and reinforce them is the most hidden and persuading of all these forces and even the neonate is not free of it. It's education if you want it, propaganda if you don't. Religion if it's good, superstition if it's bad. Statesmanship if it's big, politics if it's small. Devotion if it's right, prejudice if it's wrong. Why are people so worried because forces of the kind they always depend on are influencing them on the sly? If Freud ever did anything for human civilization, it was to show that man's motivation is ordinarily inscrutable to himself. The good teacher has got as many hidden persuaders up his sleeve as has Madison Avenue.

Contemporary Psychology, Feb., 1959, p. 46.

TWO INSTRUCTIONAL GRANT PROGRAMS

Two instructional development grant programs are now well established at Purdue and their funds are selectively awarded to faculty applicants who propose significant projects in the improvement of undergraduate teaching. The larger of the two programs, known as the Faculty Grant Program for Instructional Development and Innovation, awarded twelve grants (of up to $2500 each) in 1974 and 21 grants in 1975. The smaller program is known as the Parent's Fund for Instructional Development. It has made nine $1000 grants each year, one within each of Purdue's nine schools.

Six criteria are stressed by both programs:

1) Improving instructional effectiveness and, thus, student learning.
2) Evaluating impact of proposed strategies and, where appropriate, refining the strategies.
3) Focussing on the more serious, including large scale, instructional problems.
4) Increasing instructional efficiency and/or controlling its costs.
5) Demonstrating techniques or approaches that can be adapted or adopted by others.
6) Developing instruction from a foundation of sound principles and resources.

In the review process, each proposal is evaluated against the six criteria. Also, the proposal evaluation panels look for evidence of departmental or school commitment to each project and they expect to find willingness to commit resources roughly equalling the amount of the requested grant. The panels, incidentally, consist of 3 to 5 faculty members who are broadly representative of all schools of the University.

Funds for these programs are provided by the University Development Office and administrative responsibility is invested in the Office of Instructional Services, a unit within the Measurement and Research Center. Inquiries concerning either program can be directed to OIS (ENAD Room 402). Since OIS serves principally to organize and oversee the two programs, not to decide the merits of individual proposals, it is free to work with individual faculty in developing proposal ideas and in refining them for formal submission. Faculty are encouraged to use this help, as needed.

Formal announcements of the programs' deadlines and application procedures can be expected late in the first semester (for Faculty Grants) and near the middle of the second semester (for Parents' Fund).

* *

The success of language in conveying information is vastly overrated, especially in learned circles.

Alfred North Whitehead

TWO LOCAL TEACHING PROJECTS

The two instructional project abstracts that follow were selected from among thirty-nine that appear in the CIC publication, Development and Experiment in College Teaching (Report No. 11, Spring 1975), a publication devoted to new teaching approaches at CIC Universities, i.e., the "Big Ten", plus Chicago. These two abstracts represent two of many notable efforts being made by Purdue faculty members.

PSI in a Child Psychology Course

Each semester Psychology 235, an introductory child psychology course, serves about 250 students with diverse undergraduate majors at Purdue University. A modification of the Keller Personalized System of Instruction (PSI) is utilized to provide an individualized, self-paced teaching format.

Following a course manual written by the instructor, students independently study each course unit. When a student feels he/she has mastered the unit, he/she requests a quiz which is created when the student randomly selects five question cards from a shuffled deck constituting the pool of questions available for that unit. As many as three alternate-form quiz attempts are permitted for each unit without penalty. Quiz feedback and remediation are immediately provided by "proctors," previous students of the course selected on the basis of their excellent performance, maturity, and reliability. Demonstration of unit mastery results in points which then accumulate to determine part of each student's grade.

Points are also derived from scores received on two summative hour exams and for completion of extra projects. Projects allow students to pursue topics of special interest in greater depth. They may include written or oral reports on films, tapes, outside readings or field observation. A "mid-semester" exam is initially offered as soon as about 20 students indicate their readiness; it is then available in an alternate form at the semester mid-point. The final examination is given two weeks early for those who choose to take it and in alternate form at the end of the semester. If they wish, students may take either major examination twice and record only the higher score of the two. To discourage procrastination, students must attempt the mid-semester exam no later than mid-semester and the final no later than the end of the semester. In addition to individually paced study sessions, the class meets as a group once each week for group presentation, films, or demonstrations.

The course is rated highly by students and is widely preferred over more traditional lecture methods. The PSI method is also more effective that the lecture in promoting retention and decreasing test anxiety during examinations.

For additional information, write Professor David A. Santogrossi, Department of Psychological Sciences, Purdue University.

INSITE: A Simulation of Archeological Site Excavation

INSITE is a computer-based simulation of an archeological excavation developed at Purdue University to resolve several critical problems in teaching anthropologically-oriented archeology. While the structuring of archeological research has become more complex, involving more explicit research designs, sophisticated sampling strategies, and computer assisted data analysis, the lack of adequate training facilities in many programs hampers field training. Since the problem formulation stage of archeological fieldwork must begin at a lower level of training to accommodate the complexity of archeological research, simulations compensate very well for the lack of training facilities.

INSITE involves the simulated excavation of an artificially created archeological site by teams of students. Through the experience of actually carrying on an "independent" research project, students learn the interconnection of the various phases of archeological fieldwork. The "site" consists of a series of maps including land contour, architectural features, burials, and workshop areas. The information regarding the site is divided into bits of data representing natural excavation units (either stratigraphic layers or arbitrary levels). Data on counts of artifacts, stratigraphic information, mortuary patterns, and ecological data are stored on magnetic tape for retrieval by the student teams.

The simulation is run in a series of sessions: surface collection and test excavation; larger scale excavation and deep sounding. Since each team samples only 15-20% of the total site information, each must decide on an excavation strategy. After a sampling strategy is determined, additional information is acquired by submitting the strategy as a job request using the Purdue batch processing system. Upon receiving the requested information, each team must then modify its strategy of excavation for the following session. Thus, the research flexibility is analogous to the actual conduct of site excavation. In addition to the data stored in computer files, each team receives Xerox copies of the maps and architectural information correlating with the units chosen.

Upon completing the simulation with the exhaustion of the team's supply of excavation units, a final site report is prepared. Each team member specializes in analyzing one category of data, preparing one section of the final report. In addition to the written report, all teams prepare brief oral reports which are presented in a single session. In this way, each team can debate the effectiveness of approaches used by other teams. After all discussion ceases, the overall structure of the site is revealed and students may evaluate their methods.

For additional information, contact Professor Robert E. Fry, Department of Sociology and Anthropology, Purdue University.

TRENDS IN LEARNING AND TEACHING

A memo series for Purdue's teaching faculty, issued twice each semester by the Office of Instructional Services, Measurement and Research Center, ENAD 402.

Volume 1

Number 2

December 1975

Trends in Learning & Teaching

A memo series for Purdue's teaching faculty, issued twice each semester.

INSTRUCTIONAL GRANT PROPOSAL DEADLINE

In the first (October, 1975) issue of this memo, a page 3 article briefly describes Purdue's two instructional grant programs. By now, faculty should have received by mail this year's formal announcement and guidelines for one of those programs, The Faculty Grant Program for Instructional Improvement and Innovation.

Please note that the deadline for receipt of proposals is Friday, January 16, 1976. Faculty who have questions about the program or who desire help in the development of proposals may call Dr. W. F. Seibert at 2207.

A FILE FOLDER PRINTING ERROR

We knew it would happen, but could not define "it" until too late. In the file folder that you received with your first issue of this memo, a mistake in the printing has been found. Please note that the contents printed on the inside of the folder have reversed two of the headings. The contents headed "Audio-Visual Production Department, continued" should be "Audio-Visual Center, continued" and vice versa.

We regret the inconvenience that this error may have caused.

MORE ON CAFETERIA

The first memo in this series promised further articles on the CAFETERIA instructional evaluation system and we have chosen to discuss here some evaluation issues and other questions concerning its results, their use, and their interpretation. First, as a brief review, recall that CAFETERIA was designed expressly to serve the individual teacher and to help teachers monitor, modify, and improve their teaching. Needless to say, results such as those from CAFETERIA are just one of many kinds of information that can prove helpful. But since CAFETERIA's reports are based on students' statements and judgments, they are more practical to secure and the results more accessible than most alternatives - e.g., peer evaluations, administrator evaluations, or alumni reports and evaluations. Accessibility and practicality could be vices, not virtues, if students' judgments and perceptions were generally or seriously in error, but evidence indicates that they typically are not. Instead, students, with their considerable experience as educational "consumers", are capable of providing reliable and discriminating judgments of most teaching. So, while students' judgments are at times influenced by extraneous factors, they are no more vulnerable than the alternatives and they often appear less vulnerable.

CAFETERIA is most unique in its ability to satisfy teachers' interests in evaluating many varieties or approaches to teaching. Unlike traditional evaluation schemes, CAFETERIA makes few assumptions about a teacher's style or course plan. It seeks to accommodate a wide range of styles. Many teachers employ lecture or discussion methods, but increasing numbers also use project-based methods, individually paced instruction, mastery learning, team teaching, and other more or less novel approaches. Thus, in developing CAFETERIA, special effort was made to provide for evaluations that would be sensitive to these, as well as to more customary forms. Other important differences among courses can be associated with factors such as size of the course enrollment, the academic discipline, students' characteristics, course level, and the variety(ies) of learning desired. These are recognized both in the general design of CAFETERIA and in specific evaluation items that are included in the current 200-item catalog from which participating teachers make their choices.

Evaluation questions, issues, and the direction of CAFETERIA's future development are a continuing concern. One of the largest is a concern with balance in the use of CAFETERIA's results. Primarily, the results are meant to help teachers improve; nevertheless, they are rather frequently used in personnel review, i.e., when teachers <u>voluntarily</u> submit evaluation reports to departmental promotion, tenure, or salary review groups. For several reasons, it is hoped that this use will remain distinctly secondary. First, diagnostic instruments, like CAFETERIA, have different qualities than do instruments that seek only to arrange "examinees" on some single and general dimension of quality or "goodness". Diagnosis is a process of locating weaknesses as well as strengths, while general evaluation is a process for arranging participants along a continuum of quality. A related problem arises from the fact (or assumption) that teachers who plan to submit results for departmental review will be less likely to search out areas of instructional weakness. To identify instructional weaknesses may be a step toward later improvement, but a step away from favorable departmental consideration.

continued page 2

CAFETERIA, continued from page 1

Another set of questions concerns the methods of administering CAFETERIA to students. In general, conditions should be established that will bring forth students' thoughtful and candid judgments. For this reason, faculty are encouraged to consider doing the following: involve students, if possible, in selecting at least some of the evaluation items; choose a time of administration when attendance is high, but when students are not distracted or emotionally affected by such things as a major exam or a course deadline; set a climate that invites candor and that guarantees student anonymity. (It may be possible either to have a class member take responsibility for the administration or to trade such duties with a colleague.) Finally, there are interesting arguments for administering evaluations early enough in a semester that time remains for effecting changes; students will be more likely to evaluate carefully if they can personally benefit from changes that a teacher adopts.

The interpretation of diagnostic results is usually done by the teacher involved; however, help is available to those who request it ('phone 2207). Such help is partly a matter of benefiting from the ideas of a disinterested second party, but it is also a chance to use the CAFETERIA staff's experience in its reviews of many earlier results.

To conclude, it is hoped that all faculty will regard instructional evaluation as a non-threatening opportunity to monitor and improve their teaching. It is a service that is almost constantly available in whatever volume the faculty requires. It also is a developing service, one that can benefit from the suggestions of its users and for this reason, your suggestions are sincerely invited.

****

The true critic...is the man who becomes your personal enemy on the sole provocation of a bad performance, and will only be appeased by good performance.
George Bernard Shaw

****

TV RECORDINGS FOR TEACHING IMPROVEMENT: A NEW SERVICE

In recent years, portable television recording equipment has found wide use in a variety of teaching improvement projects and faculty generally find these uses to be practical and productive ways to review and analyze their teaching styles. With that in mind, a Sony "Port-a-pac" TV system has been acquired by the Measurement and Research Center for loan to interested faculty. The "Port-a-pac" consists of a small television camera with a zoom lens and microphone, a 15-pound self-powered (or externally powered) television tape recorder, and a television receiver-monitor for viewing the recorded materials.

The equipment is designed for simplicity, so interested users can learn its operation readily. After a few minutes' orientation, most are able to master the basic operations required to "make television" of classroom lectures, discussions or similar sessions. The recordings can then be viewed later, repeatedly, if you like, and analyzed or critiqued. Under most conditions, the quality of recorded picture and sound is more than satisfactory.

This system is available to faculty on a "first come-first served" basis and it is suggested that interested faculty may want to use it initially for some exploratory trials. Then, if they choose to pursue the work further and perhaps to develop sophisticated ways to use the system for instructional improvement, the staff of MRC and the Office of Instructional Services is available to assist.

Interested faculty may arrange for loan of the "Port-a-pac" by calling Mr. John Van Horn, 'phone 2207.

****

O wad some power the giftie gie us,
To see ourse1s as ithers see us!
It wad frae mony a blunder free us,
An' foolish notion,
What airs in dress an' gait wad lea'e us,
An' ev'n devotion!
Robert Burns

****

Knowledge is of two kinds: we know a subject ourselves, or we know where we can find information upon it.
Samuel Johnson

****

DEAN A. A. POTTER

Few institutions have the good fortune to retain a living legend whose reputation spans several generations and who, at age 93, is still as committed and articulate as the best of younger colleagues. But Purdue has Dean A. A. Potter, a legend and a pillar without equal. He was invited and he agreed to present here his reflections on teaching.

For those too young to know "The Dean", he is a native of Vilna, Russia, and he emigrated alone to America as a very young man. He is a 1903 graduate of M.I.T., and he served not only as Purdue's Dean of Engineering for many years and interim President during 1945-46, but in many other roles and assignments. Outside the University, Dean Potter served on countless boards and councils in business, government, and professional societies; these include the policy board appointed by President Truman to establish the National Science Foundation. Officially, Dean Potter retired a quarter century ago, but not before meeting and serving every American President from Woodrow Wilson through Dwight Eisenhower. (Coolidge, he says, "....did not say much, because he had little to say.") But, although retired, he still spends mornings in his campus office.

Dean Potter's contributions encompass science and academic policy, research and the publication of 300 papers, employment as an engineer in industry, technical consulting, and a half century in teaching. His other enjoyments have always included music, literature, languages, and theatre.

We think you will find his thoughts on "Effective Teaching" provocative and interesting.

Editor

EFFECTIVE TEACHING By: Andrey A. Potter, Dean Emeritus of Engineering, Purdue University

As a teacher and dean of Engineering for 48 years, 33 at Purdue and 15 at Kansas State University, I selected and trained many teachers. I found that the best teachers have to their credit the following attributes:

1. Knowledge of the subject he or she teaches.
2. Effective presentation. This heading includes such qualities as the ability to hold the student's interest, proper coordination of the theoretical and practical, power of expression, enthusiasm, human approach, emphasis of important points, persistence with the slow student, and thoroughness.
3. Desirable traits. In this category should be included strong character, superior ability, good attitude, balanced philosophy of life, alertness, broad-mindedness, courtesy, friendliness, diplomacy, energy, proper ethics, fairmindedness, gentleness, industriousness, kindness, neatness, sense of humor, sincerity, thoroughness, tact and broad tolerance.
4. High culture and superior general education.
5. Effectiveness in class management.
6. Practical experience.

Specifications for an effective teacher should also include the following: The teacher must first of all be a person of character. A person who has low ethical standards has no place in the American educational system. A teacher who condones dishonesty, deceit and trickery, or who evades the laws of the land, even in spirit, is neither a good teacher nor a good citizen. A teacher who lacks sincerity or who bluffs or exaggerates is a detriment to any educational institution. A drunkard, a libertine or a liar must not be tolerated in the teaching profession where the well-being of young people is involved. A teacher who is absolutely moral, clean, just, dependable, accurate, mentally honest, industrious, conscientious, and courageous is the type of person who is in a position to instill sound character and an appreciation of right living in his students. I am puzzled by the code of ethics of certain college teachers. In most professions the practitioner guards most jealously the reputation of others in his specialty . University teachers have been known to criticize their own colleagues and even to run down the competence of teachers before their students.

Unbounded enthusiasm is a major characteristic of the good teacher. His enthusiasm radiates to those about him in the sparkle of his eye and the vibrancy of his voice; students are fascinated by his general animation and by his interest in the subject he teaches. When a teacher loses his enthusiasm and his ability to inspire his students with an eagerness to learn and an interest in the search for knowledge he should retire from teaching. Teachers will do well to study the methods of the successful coaches of our colleges, and they will realize that character, personality, enthusiasm and superior teaching techniques on the part of such coaches are largely responsible for the victories of athletic teams.

Sarcasm, conceit, snobbishness, selfishness, profanity, antagonism, or intolerance are traits which characterize the poor teacher. A person who has any such traits or is careless about his dress or appearance has no place on a university faculty.

No teacher should be tolerated if he is merely a hireling and a timeserver; he must be bubbling over with enthusiasm for the subject he teaches if he is to inspire his students to become scholars and thinkers. Neither is there a place for a teacher who is impatient, who lacks self-control, who is non-cooperative or disloyal. A teacher who cannot be loyal to the institution he serves and to its policies should seek opportunities elsewhere. No self-respecting individual accepts compensation from an organization to which he cannot devote himself loyally and unselfishly.

The master teacher is always agreeable, courteous, considerate, kind, patient, diplomatic, optimistic, tactful, and inspiring. He commands the respect of his students and is in a position to develop in them habits of courage, straight thinking, sound reasoning, industry, and a definite interest in the intellectual and the aesthetic. It was Goethe who said, "We learn only from those we love."

The good teacher is a clear thinker and bases his statements upon facts. He thinks in terms of principles and is not merely an imitator of others. He has the ability to analyze and synthesize the facts involved in any problem and arrives at solutions objectively without prejudice. He has cultivated in himself intellectual curiosity and habits of study, and is able to instill these qualities in his students.

The effective teacher is well versed in his mother tongue and has the ability to express himself clearly and correctly. He can talk audibly enough to be easily heard and his enunciation is such as to be easily understood. He is constantly enlarging his vocabulary through his reading and encourages his students to do likewise.

The good teacher is a master of the subject he teaches and is able to draw out from his students the best that is in them. One whose education has been limited only to the subjects that he is trying to impart is unfit to occupy a post at an institution of higher learning; the teacher must broaden his vision through advanced study and experience before he attempts to teach, particularly at the institution from which he received his bachelor's degree. Graduate study, engineering experience, reading, travel, and a keen interest in young people and in their problems are essential prerequisites. A person who has never worked for pay in a shop, plant or on a construction job, and who has never rubbed elbows with workers, skilled as well as unskilled, lacks essential experience for engineering teaching. The judgment of values which the engineer must possess can best be acquired through experience with men, things and ideas. At the same time it must be remembered that knowing only the facts to be taught does not necessarily make a good teacher. Those in executive positions in colleges and universities have had sad experiences when they added to their staff engineers of wide practical experience, who knew nothing about teaching, who did not realize that teaching is an art, and who merely imparted knowledge of a very specialized type.

To be effective one must teach people and not subjects. The teacher must know and must have a sympathetic interest in each student. He must base his instruction upon the preparation and aptitude of each of his students and not upon a theoretical median.

continued page 4

EFFECTIVE TEACHING, continued from page 3

The good teacher aids his students in learning how to study, how to plan their work, and how to reduce lost motion. He develops the students' proper mental attitude toward their studies and surroundings. He is constantly pointing out applications of the problems under consideration. He is aiding his students to differentiate between facts and opinions, and impresses upon them the necessity of developing an analytical attitude and a social consciousness.

Research in an educational institution should have as its main objective the improvement of its principal product, the student. Research, either in methods of teaching or in the specialty of the teacher, is definitely helpful to the teacher in keeping up to date and in increasing the inspirational value of his teaching. In fact every good teacher is necessarily a scholar. Professor Hardy Cross of Yale stated that "A teacher, who is constantly trying to master his field, almost invariably produces research, books, articles and addresses. He wants to know, not only to be a good research man but in order that he may teach well." While I am the last person to recommend that the advancement of teachers be based solely upon their contributions through research, the fact remains that most teachers who carry on no research, who do no writing and who are unknown even in their own field have little inspirational value and are usually found to lack initiative and ability to improve themselves or others. Bridge, golf, household tinkering and poor planning of one's time account for the relatively insignificant productivity of the majority of college teachers. Teachers who are receiving little advancement will find upon careful analysis that they through their own laziness are definitely in a rut; a rut, it should be remembered, differs from a grave only in dimensions and outlets.

In the report of the Investigation of Engineering Education, conducted by the Society for the Promotion of Engineering Education, the conclusions were "that the real problem before us is that of securing the right type of teachers for engineering faculties. On it hangs the satisfactory solution of our most serious educational problems. Engineering education will never be better than its teachers." There is always a scarcity of inspiring teachers who are known to be masters in their own fields. However the best colleges are not interested in teachers who merely meet their classes, who show no personal or sympathetic interest in their students, who contribute nothing to the advancement of their profession, and who are unwilling to prepare themselves by constant study to be recognized as experts and scholars.

A college or university is the top story of the educational structure; it is made up of a number of persons in one body, something like the guild of old, but whose main interests are concerned with creating, imparting, preserving and acquiring knowledge. In the true university all are learners--teachers and pupils alike are scholars. There should be very little gap between the student and the teacher and absolutely no bridge between teachers of different ranks. Academic titles, such as instructor, professor, dean and president are intended merely to be rewards for scholarship and effectiveness of individuals as teachers and/or creators of knowledge. Such titles, however, are not intended to involve military line authority. Instructor, assistant professor, professor, dean and president are all colleagues who should cooperate in the interest of the learner and for the advancement of scholarship.

Teaching is an idealistic profession, like music or painting. The teacher, like the artist, does his best if he works in an environment which is pleasant and free from organizational worries. The higher the rank the teacher holds, the greater is his responsibility for true scholarship and for the guidance of others. The young instructor teaches routine and elementary subjects. As he gains mastery of his specialty and proves himself able to inspire his student and to create new knowledge or to develop better teaching methods through his own scholarship, he is advanced in rank and responsibility. The title professor indicates the highest honor the university can award to any teacher; it stamps him as an inspiring guide of his students and as an authority in his own special field.

The head of the department is a person who, in addition to his standing as a teacher and authority in his specialty, is expected to have the ability to inspire his younger associates to greater endeavor. A department head who treats his colleagues as his subordinates or who feels that he has to bear down upon the teachers he is expected to develop, stimulate and encourage is more fitted to be a boss of a construction gang than a professor. A dean, like a department head, should be a teacher and investigator who continues his scholarly endeavors and not a mere administrative officer. He is expected to interpret the needs of his staff to the president and trustees of his university, and must be the guardian of the standards and prestige of the school entrusted to his care. His major task, however, is that of a catalyst who stimulates, inspires and encourages his colleagues and students in their scholarly endeavors.

Finally, it must be remembered that the compensation of the effective teacher is not represented by his paycheck; much of it comes in the reward of work well done in an atmosphere which is conducive to high ethical standards and productive scholarship. The master teacher loves his work and derives satisfaction from the accomplishments of his former students. Furthermore, those of us who are on the staff of higher educational institutions enjoy a social status which money cannot buy; the families of college faculties have cultural and educational advantages which those in business or in other professions can purchase only at a high price, if at all. A fair appraisal of the life and work of the college teacher leads to the conclusion that he is most fortunate in the rewards and satisfaction of life which his calling afford him.

****

A FINAL NOTE: In writing, the "he/she", "his/hers" problem is of recent origin and not yet mastered, so far as we know, by any writer. Readers sensitive to this might note that Dean Potter's daughter is among those whom he addresses here; she holds degrees from Vassar and Johns Hopkins and is a Purdue Professor.

Editor

TRENDS IN LEARNING AND TEACHING

A memo series for Purdue's teaching faculty, issued twice each semester by the Office of Instructional Services, Measurement and Research Center, ENAD 402.

Volume 1

Number 3

March 1976

Trends in Learning & Teaching

A memo series for Purdue's teaching faculty, issued twice each semester.

SPRING SEMESTER CAFETERIA SERVICE: AN INVITATION TO ALL FACULTY

In past semesters, the Measurement and Research Center (MRC) has employed special mailed notices - invitations to the faculty to invite all to participate in the CAFETERIA instructional evaluation service. This semester, the invitation is being included here. Necessary information is presented below.

Any faculty member, teaching assistant, or faculty group may arrange to receive the service by contacting MRC ('phone 2207) and requesting catalog(s) and item selection form(s). Following receipt of your completed item selection form(s), the MRC staff can assure that your request(s) will be processed and ready for classroom administration within five working days of the time when the request is made. During slack times, such requests are typically filled more promptly, often within a day or two.

Faculty are again reminded that CAFETERIA is a confidential service to those who participate and it is designed primarily to provide diagnostic information that helps to identify instructional weaknesses, as well as strengths and this in order to guide course revisions or other instructional improvements. Its contents also are meant to be adaptable across a wide range of course types and teaching styles.

Anyone with questions about the service or its availability is encouraged to contact MRC.

****

All great ideas and all great actors on the stage of history were abnormal. Was Beethoven normal? Or Michelangelo? Da Vinci, Socrates, or Mohammed?... Treat gently the abnormal; he may carry some subtle talent under the cloud of his peculiarities.

Dagobert D. Runes

****

INSTRUCTIONAL GRANT PROGRAM PROPOSAL DEADLINE SET FOR MAY 21.

In former years, faculty have been invited to submit instructional grant proposals to the Parents' Fund program. That program is being continued this year but under a new name, the President's Faculty Award Program. Formal announcements and guidelines for the newly named program will be sent to faculty shortly. These announcements - guidelines will indicate that the proposal deadline will be May 21. Grant application procedures and program criteria will be set forth also. As in the past, nine project grants will be available, one within each of Purdue's nine schools. The size of each grant, as in the past, will be $1000.

The President's Faculty Award Program is administered by the Office of Instructional Services, Measurement and Research Center (phone 2207). Interested faculty are encouraged to contact OIS, if there are questions.

ROLE MODELS FOR WOMEN SCIENCE UNDERGRADUATES

"Women and the Science Disciplines", an alternate semester course, now in its third year, was initially implemented as part of a school-wide experimental two year program to reduce the excessive attrition rate of women undergraduates in the science disciplines by providing a series of viable role models. The course, which is team-taught, uses a seminar format consecutively covering each of the seven discipline areas in the School of Science. Thus, after the initial weeks spent in exploring the nature of the relationship between science and society and the historical role and contributions that women have made to the sciences, a typical week will include an overview lecture by a faculty member of the discipline under discussion, followed by a presentation given by a successful woman practitioner of that discipline, discussing her professional activities and as much of her personal history and lifestyle as is appropriate. Some class time is also devoted to discussion of employment prospects, the advantages and disadvantages of graduate education, beneficial experiences available to undergraduates such as NSF summer internships and research and teaching assistantships and the "nuts and bolts" of how these things can be accomplished.

Analyses of attrition data and student evaluations have shown that the course is effective in helping to reduce the attrition rate. General faculty response to the course has also been extremely favorable, as demonstrated by the willingness of the faculty to participate as lecturers and resource persons.

continued page 2

ROLE MODELS, *continued from page 1*

As of Spring, 1976, "Women and the Science Disciplines" also became one of the radio courses offered to the general public through Continuing Education and is thus aired weekly on Purdue's radio station, WBAA.

For further information, contact Lynne Harrington Brown, School of Science, Mathematical Sciences Building, Campus.

****

The problem of a coherent civilization is the problem of living with ignorance and not being frustrated by it.

J. Robert Oppenheimer

****

PSI, AN OPTIONAL MODE OF INSTRUCTION IN A STATISTICS COURSE

Students in Statistics 213 may elect to follow a self-paced instructional format developed as a modified Keller personalized system of instruction. Statistics 213 is a first course which teaches basic probability and statistical reasoning with emphasis on mathematical models and problem formulation. A major course objective is to prepare students for later quantitative studies. Among the 500 students taking this course each semester there is a variety of mathematical preparation and ability.

The instructor has divided the course material into 15 coherent units and written a study guide relating each unit to the textbook where additional background information is needed, gives reading and problem assignments from the textbook, delineates unit objectives, offers solutions to selected textbook problems and contains a sample quiz. Students study independently but have available a cadre of undergraduate "tutors" who staff a resource room throughout each day. Students demonstrate mastery of each unit by taking a tutor-administered quiz. Quizzes are graded in tutorial fashion with student and tutor discussing any errors. Three examinations serve as review units. A series of computer simulation projects are used to add empirical credence to abstract concepts. Quizzes, examinations and simulation projects each contribute points to a cumulative score for grading purposes.

Both quizzes and examinations may be repeated to improve mastery. Alternate forms of all test material maintain the integrity of the testing process. A minimal rate of student progress is assured by deadlines set for the completion of each of the three examinations. Audio tapes, prepared by the instructor, guide students in the analysis of each computer simulation.

Student attitude toward the course is positive. Since students are self-selected for this mode of instruction, it is difficult to make comparisons with the course taught traditionally in small sections. Students of widely differing abilities have been equally successful in meeting course objectives. Students consistently request that subsequent courses be offered in the same format.

Additional information, sample materials and results of a pending study of student success in subsequent courses may be obtained by contacting Professor James Yackel, Department of Statistics, Campus.

An enemy is willing to devote a vast amount of time and brain power to ferreting out errors both large and small, and this without any compensation. The trouble is that really capable enemies are scarce, and most of them are only ordinary. Another trouble with enemies is that they sometimes develop into friends and lose a good deal of their zeal. It was in this way that the writer lost his three best enemies.

Georg von Békésy

Cafeteria Deadline

Faculty are asked to submit their requests for CAFETERIA instructor and course appraisal service to the Measurement and Research Center (MRC) by April 16. Requests received by that date can then be processed and the individually tailored rating forms can be supplied to faculty before the end of the semester. For information, catalogs, and item selection forms, please call MRC, 749-2207, or come to Engineering Administration, Room 402.

April 16

TRENDS IN LEARNING AND TEACHING

A memo series for Purdue's teaching faculty, issued twice each semester by the Office of Instructional Services, Measurement and Research Center, ENAD 402.

CUT OR TEAR ALONG THE CREASE AND RETURN THE COMPLETED QUESTIONNAIRE VIA CAMPUS MAIL

A "Straw Poll" on Instructional Evaluation

The CAFETERIA instructional evaluation service has seen expanding use since its inauguration in 1972. It has been refined in several ways during this time and yet three basic and broad questions about the present system-and-service still remain: 1) how satisfactory is it for the faculty, 2) how is it being used, and 3) what can be done to improve the service.

Please take a few minutes to consider and respond to the questions below. Then, fold this sheet, staple it, and drop it in campus mail.

Your answers will be useful, even if you have not used CAFETERIA's services.

---

1. Your academic rank:
   - _____ professor
   - _____ associate professor
   - _____ assistant professor
   - _____ instructor
   - _____ graduate teaching assistant
   - _____ other

2. Your school of the University:
   - _____ Agriculture
   - _____ Engineering
   - _____ Home Economics
   - _____ HSSE
   - _____ Industrial Management
   - _____ Pharmacy
   - _____ Science
   - _____ Technology
   - _____ Veterinary Science and Medicine

3. About how many years of teaching experience have you had?
   _____ years

4. Have you used the CAFETERIA instructional evaluation service?
   - _____ yes, regularly
   - _____ yes, some
   - _____ no, never

5. If yes to question 4, what thing(s) prompted you to use CAFETERIA? _____

6. If no to question 4, please indicate your reasons AND please indicate below how the system might be changed to meet your needs or reservations. _____

7. What are your feelings about (a) a policy that would require all faculty to be evaluated by students, at regular intervals (e.g., once each year)? _____

and (b) the use of students' instructor ratings as one factor or element in faculty personnel decisions (e.g., who shall be promoted, receive tenure, receive salary increases)? _____

(The following questions apply only to those who have used CAFETERIA services.)

8. When you have used CAFETERIA, have you found the results useful? (Please explain) _____

_____

_____

_____

9. Are there changes or extensions of CAFETERIA that would make the service more useful to you? (Please explain)

_____

_____

_____

10. Specifically, do you find CAFETERIA's item catalog broad and inclusive enough to meet your needs? (If you like, you may suggest here new items that you would find useful.) _____

_____

_____

_____

11. When you have received CAFETERIA results, how have you used them (e.g., to develop course revisions, to build a case for your promotion or salary increase, to seek interpretations-advice from a colleague)? ____

_____

_____

_____

12. Help in interpreting evaluation reports is available from staff consultants at the Measurement and Research Center. Have you made use of these services?

_____ yes, and they were helpful.

_____ yes, but they were not especially helpful.

_____ no. (Please elaborate) _____

_____

_____

_____

CAMPUS MAIL

TO: Measurement and Research Center
ENAD 402

CAMPUS MAIL

Volume 2

Number 1

December 1976

Trends in Learning & Teaching

A memo series for Purdue's teaching faculty, issued once each semester.

A NEW CMS SERVICE (COURSE MANAGEMENT SYSTEM) NOW AVAILABLE

For some time it has seemed that faculty could be well served by a system which helped in organizing, summarizing, and maintaining records of their students' progress. The need seemed especially great in large enrollment, multiple section courses and in those that employed flexible or self-paced forms of instruction. For example, in multiple section courses, which may also involve various indications of student progress (quiz scores, lab performance scores, examinations, special projects, attendance records, etc.), a large quantity of student performance information must be collected and maintained; ideally, this information should be current throughout each semester. If current, then the instructor(s) in charge can identify students who are not progressing well, can report regularly to any student concerning his/her status and prospects in the course, and can monitor such things as the status or progress of student groups enrolled in various course sections.

During the last several months the Data Analysis Section of the Measurement and Research Center has developed a CMS system and begun providing its services to a number of fall semester courses. The CMS reports that are provided to faculty include such information as updated (cumulated) student scores; records of individual scores on examinations, ordered alphabetically, by score value, or by students' SS number; and several optional or specially processed reports.

Faculty interested in discussing CMS services are invited to contact Mr. John Van Horn, MRC, 'phone 2207.

TILT PUBLICATION SCHEDULE

At the outset, it was anticipated that TILT would be issued twice each semester. Experience during the first year of publication (1975-76), similar experience reported by other institutions, and the ever-present need to economize all suggest that this be changed. TILT is now scheduled for one issue per semester.

Editor

FYI

The November 15 issue of The Chronicle of Higher Education reports on a recent conference which dealt with self-paced instruction. Readers may find the report of interest, since it discusses methods which have been increasingly used and accepted in recent years. While not an epidemic, at least a trend is emerging and good results are being regularly reported. (Two of the many local instances of self-paced instruction are briefly described elsewhere in this issue.)

The adoption of self-paced instruction can originate from several sources, but those that are often cited include the need to accommodate to a diversity of students' backgrounds and abilities, limited resources and logistical problems encountered in more traditional instructional arrangements, dissatisfaction with the levels of mastery or comprehension which students may customarily achieve, a desire for higher levels of student motivation, and a greater efficiency in the instructional process.

What do the higher schools, those to which the great middle class of the country sends its children, teach, over and above the instruction given in the primary schools? There is a little more reading and writing of English. But, for all that, every one knows that it is a rare thing to find a boy of the middle or upper classes who can read aloud decently, or who can put his thoughts on paper in clear and grammatical (to say nothing of good or elegant) language. The "ciphering" of the lower schools expands into elementary mathematics in the higher; into arithmetic, with a little algebra, a little Euclid. But I doubt if one boy in five hundred has ever heard the explanation of a rule of arithmetic, or knows his Euclid otherwise than by rote.

Thomas Huxley, Science and Education, Essays, 1904

A SELF-PACED INSTRUCTIONAL APPROACH FOR A DYNAMIC SYSTEMS AND CONTROLS LABORATORY

The School of Mechanical Engineering moved into modernized teaching laboratory facilities at the end of the 1975-1976 academic year. The laboratory sections of four required ME courses (Fluid Mechanics, Heat Transfer, Dynamic Systems and Measurements, and Dynamic Systems and Controls) share one floor without any permanent separating walls. Each course serves approximately 200 students per year. A self-paced instructional approach has been developed for the senior level Dynamic Systems and Controls laboratory to promote efficient utilization of space while improving the quality of the student laboratory experience.

Important goals include allowing students to explore the properties of dynamic control systems at their own pace and emphasizing a realistic, creative laboratory situation. Under the self-paced format, the student is assured convenient access to available instrumentation or equipment on an open laboratory time basis (60 hours/week).

The laboratory experience is divided into 21 self-paced instructional (SPI) blocks, with eighteen required for minimum credit. A clearly stated hierarchy showing any block prerequisites facilitates student learning. After studying the instructional materials for a block, the student must demonstrate mastery of behavioral objectives. Assessment tasks may require oral or written responses; in either case students are allowed to defend their answers orally. If the student does not fully meet the block objectives, the laboratory instructor offers guidance; after additional study, the student may recycle until mastery is demonstrated with no stigma attached.

Thirteen orientation blocks cover necessary skills for equipment operation or introduce basic laboratory concepts. Five have audio-tutorial tape cassette materials to augment written instructional materials. The other eight blocks are hardware-oriented experiments; in three cases the student must submit in writing a proposed experiment procedure of his or her own selection and in three others a written laboratory report is required.

This self-paced instructional approach has been used in a twenty student pilot section with very encouraging results, based upon subjective views of the three laboratory instructors involved and on CAFETERIA questionnaire evaluations by students. Because of this preliminary success, all Dynamic Systems and Controls laboratory divisions are being combined into a single self-paced program.

For additional information contact Professor E. Dawson Ward, School of Mechanical Engineering, ME Building, or 'phone 39850.

OFFICE OF INSTRUCTIONAL MATERIALS DEVELOPMENT (OIMD)

The Office of Instructional Materials Development, headed by Richard O. Forsythe, has been established to provide a single point of contact within the University for faculty members and departments interested in developing new teaching strategies and materials. A number of agencies (i.e., Measurement and Research Center, Library, Audio-Visual Center, Computing Center, Radio Station WBAA, Telecommunication Center, and Audio-Visual Production Department) provide services to assist the improvement of instruction and learning. The OIMD was implemented to provide a single point of contact for those departments and individual faculty members who are interested in major efforts to improve the efficiency and effectiveness of instruction through the incorporation of media and other University learning resources. Such efforts often require resources from a variety of design, evaluation, and production departments. The OIMD can provide the coordination of efforts necessary to focus all of the available services upon the solution of major instructional development efforts.

Three current projects of OIMD include development of an audio-print format course in Personal Finance (MGMT 442), scheduled for availability in the fall of 1977; a three-to-five year course development project for the basic speech course (Com 114), Introduction to Speech Communication, and a companion course for its teaching assistants; and a cooperative project with the Libraries-Audio-Visual Center to establish an Independent Study Center from which students may receive access to a variety of self-paced instructional materials.

Faculty interested in the services of OIMD are invited to contact Professor Forsythe at 2381 (or Room B-10 Stewart Center).

Many of my friends, men like Rachmaninoff and Kreisler, never gave a lesson to anybody. Such people were astonished at my teaching, and they asked me, "How can you do it?" To many of them, teaching is somehow a lesser activity--only concerts matter. Certainly the world looks at it in that light. Therefore, an artist gets thousands of dollars for one concert, and only a few dollars for one lesson. Furthermore, being a teacher seems to reduce his rank. Now, I find that deplorable. Money, large audiences, standing ovations--they're all false and ridiculous criteria to a true servant of music.

Gregor Piatigorsky, Cellist
New York Times, Sept. 19, 1976

THE BIASING OF CAFETERIA'S RESULTS: A STUDY REVIEWED

Professor G. B. Ranney, a statistician at the University of Tennessee, has just completed a study into possible instructor-induced bias of CAFETERIA ratings, as a consequence of the item selection process. The University of Tennessee adopted CAFETERIA last year and, as at Purdue, it is used for faculty self-improvement and, in some departments, may be reviewed by administrators as well. These two evaluation purposes, self-improvement and administrative review, are not always complementary and thus, as Ranney suggests, "there may be some pressure upon the individual faculty member to attempt to obtain favorable responses from students rather than to obtain purely objective feedback for use in improving teaching."

Ranney's primary concern is directed toward the item selection process, usually considered a CAFETERIA strength. With CAFETERIA, individual instructors select up to forty items from a 200-item catalog, and these selections, along with the mandatory five-item core, are assembled by computer and printed in quantity. Since the catalog items always precede the core when printed, instructors might be tempted to create a "positive response bias" by selecting only items on which they expect high ratings. Thus, according to Ranney, the instructor-selected optional items may induce a favorable response set or bias that extends to the five-item core. If the core item responses can be manipulated in this way, CAFETERIA results would lose validity for such purposes as promotion and tenure consideration.

To determine whether CAFETERIA lends itself to such manipulation, Professor Ranney selected seven professors who teach two sections of the same course. She asked them to "rig" CAFETERIA questionnaires, first by selecting five optional catalog items on which they expected to receive high ratings, then five on which they expected low ratings. Two sets of questionnaires were thus constituted: one set "rigged" positively and the other negatively. These were administered to intact classes, so that each professor was evaluated by students in one class using the "positive" instrument and in the other class using the "negative" one. Ranney then analyzed the response distributions for each set of student questionnaires (N=487).

First, as expected, the five core items discriminated, i.e., faculty differed widely in their ratings on core items. Second, faculty were successful in choosing items (or rigging results). Student response distributions differed significantly ($p<.01$), indicating the presence of instructor-induced bias for the optional catalog items. Finally, and very much the point of this experiment, Ranney tested to determine whether such bias extends also to ratings on the core items. Response distributions were not significantly different and so Ranney concludes, "it appears that it may not be possible to bias responses to core items by the selection of favorable optional items; therefore, CAFETERIA may not lend itself to manipulation of the type considered (here)."

This study seems straightforward and even simple, in the best sense of that word. Of course, Ranney includes the obligatory cautions about generalizability of results and indicates that existing conditions dictated the study design. Yet the study and its results are highly relevant, not only to Purdue faculty but also to faculty at approximately sixty other institutions which use CAFETERIA. Perhaps because the results are relevant, those obligatory cautions must be seconded. The MRC staff created CAFETERIA as an alternative to the traditional standardized rating forms and primarily to serve diagnostic needs. That instructors may not be able to bias core ratings through the item selection process is no special warrant for the use of such ratings for annual faculty review. If responses to CAFETERIA's five-item core are to be used in considering promotion and tenure cases, the justification must still be based on a clear understanding of what students' ratings measure.

In summary, results from CAFETERIA's five-item core did not appear to suffer from the experimental efforts to rig the ratings. One possible explanation for Ranney's results can be found in her assumption that the core items measure teaching effectiveness. Assume for the moment that these core items measure student satisfaction with the instructional experience, teaching effectiveness notwithstanding. In this case, satisfaction would be a type of "halo", a disposition to rate either favorably or unfavorably. If anything, such dispositions would distort ratings received on the optional catalog items, by some more or less constant value, a prediction which may influence but not damage diagnostic interpretation. In the extreme case, the instructional weaknesses of an adored professor may be no worse than the strengths of a (much) more unfortunate colleague. Thus, comparisons among faculty may be misleading, but not necessarily for the reasons Professor Ranney investigated.

For my own part, Ranney's study has special importance because she demonstrates that the instructors who participated can indeed identify their instructional strengths and weaknesses, and that students, working independently, concur. This finding speaks directly to the diagnostic credibility of CAFETERIA, and the main purpose for which the system was designed.

James O. Derry

If you treat an individual as he is, he will stay as he is, but if you treat him as if he were what he ought to be and could be, he will become what he ought to be and could be.

Johann Wolfgang von Goethe

ROLE MODELS FOR WOMEN SCIENCE UNDERGRADUATES

"Women and the Science Disciplines" was initially implemented as part of a school-wide experimental two-year program to reduce the excessive attrition rate of women science undergraduates by providing a series of viable role models. The course, which is team-taught, uses a seminar format consecutively covering each of the seven discipline areas in the School of Science. Initial weeks are spent in exploring the nature of the relationship between science and society and the historical role and contributions that women have made to the sciences. Then a typical week will include an overview lecture by a faculty member of the discipline under discussion, and a visit from a successful woman practitioner, discussing her professional activities and as much of her personal history and lifestyle as is appropriate. Some class time is also devoted to consideration of employment prospects, the advantages and disadvantages of graduate education, beneficial experiences available to undergraduates (i.e., NSF summer internships, research and teaching assistantships), and the "nuts and bolts" of how these things can be accomplished.

The course is offered in alternate semesters and has been offered for three years. Analyses of attrition data and student evaluations have shown that the course is effective in helping to reduce the attrition rate. General faculty response to the course has been extremely favorable, as demonstrated by the willingness of the faculty to participate as lecturers and resource persons.

As of Spring 1976, "Women and the Science Disciplines" became one of the radio courses offered to the general public through Continuing Education and was aired weekly on Purdue's radio station, WBAA.

For further information, contact Lynne Harrington Brown, School of Science, Mathematical Sciences Building, Purdue, or 'phone 48326.

We know that there is no absolute knowledge, that there are only theories; but we forget this. The better educated we are, the harder we believe in axioms. I asked Einstein in Berlin once how he, a trained, drilled, teaching scientist of the worst sort, a mathematician, physicist, astronomer, had been able to make his discoveries. "How did you ever do it," I exclaimed, and he, understanding and smiling, gave the answer: "By challenging an axiom!"

Lincoln Steffens

CAFETERIA: NEWS TRAVELS FAST

Since CAFETERIA course-and-instructor evaluation services were introduced at Purdue in 1972, word concerning the system has spread fairly well to far parts of the country, and even internationally. Perhaps 50 or 55 other institutions have now adopted CAFETERIA, or some variant of it.

Also, as the word has spread, it sometimes has been garbled in transmission. The record for garbled transmission was set in late July, when the CAFETERIA offices received a commercial catalog offering to supply uniforms for food service workers -- jackets for busboys, smocks for service girls, chef's caps, the works. (The offer was declined.)

Most people are determined to hold the line against animals. Grant them the ability to make linguistic reference and they will be putting in a claim for minds and souls. The whole phyletic scale will come trooping into Heaven demanding immortality for every tadpole and hippopotamus. Better to be firm now and make it clear that man alone can use language to make reference.

Roger Brown

TRENDS IN LEARNING AND TEACHING

A memo series for Purdue's teaching faculty, issued once each semester by the Office of Instructional Services, Measurement and Research Center, ENAD 402.

Volume 2
Number 2
April 1977

Trends in Learning & Teaching

A memo series for Purdue's teaching faculty, issued once each semester.

THE USE OF CAFETERIA RESULTS IN P-S-T* REVIEW

When CAFETERIA instructional evaluation services were introduced to the Purdue faculty (in 1972), the emphasis was upon instructional diagnosis and the improvement of teaching techniques or strategies. Even though this emphasis remains, there are mounting indications that the faculty's evaluation results, or portions of them, are being submitted (by faculty) and considered (by departmental or school committees) in connection with promotion-salary-tenure reviews. When such use is made, several questions arise and a review of these questions will, we trust, be helpful in the later reviews of faculty performance.

It should first be understood that individual evaluation results are returned to individual faculty by the Measurement and Research Center; the staff of MRC does not provide these results to promotion committees, department heads, or others. Thus, when these results are used in the review of faculty performance, they are provided by individual faculty members. CAFETERIA service remains a confidential service to individuals.

In using CAFETERIA's results to make judgments about faculty performance, the largest question is one of compatibility with diagnostic purposes. To what extent will faculty use the diagnostic capabilities of CAFETERIA when they expect that these results may also influence judgments of general merit? Or, why would a teacher seek evidence of instructional weaknesses, if that evidence may lessen the chances for promotion, tenure, or increased salary?

A second question concerns the quantity of CAFETERIA results which is appropriate to provide and to use in a review of merit. There are several parts to this question: 1) Will results from the five "core" items suffice or are more items and results needed? 2) Are the results from a single course (section) sufficient or are more needed? And, 3) how many student raters are needed to provide stable, generally reliable evaluations from a course (section)?

In response to these questions, in order, it is usually conceded that the five "core" items (or, if a department chooses, a slightly expanded core) are sufficient and are preferable to the results from a larger number of predominantly diagnostic items.

When these results are merged with evaluative information from other sources, a few core items usually provide a sufficient sample of systematic student judgments. With respect to single vs. multiple course (section) results, results from several courses, preferably from several semesters, are recommended. Such a sampling provides some opportunity to judge the teacher's flexibility, resilience, and perhaps his/her changes over time, while results from one or two courses lack these advantages. Finally, with respect to the number of student respondents needed to achieve stable results, the rule of thumb is to hope for at least fifteen student respondents, but also to discount results based on fewer than ten students. One should also bear in mind that, even if the number of students is, say, 25, these should represent the majority of enrolled students; 25 responding students from a course that enrolls 50 may indicate some bias in the results.

A few words about the use of item norms are in order. Ideally, norms would compare ratings only within groups of similar or reasonably homogeneous instructors and courses, but current CAFETERIA norms do not. Instead, each course and each instructor is compared against an undiscriminated -- although usually quite large -- group of others. Thus, in reviewing the percentiles reported for an instructor, some consideration should be given to whether the course enrollment was large and perhaps unwieldy, whether the course attracts self-selected and motivated or reluctant students, the type of course and the demands it necessarily places on students, as well as the level of teaching experience the faculty member has, both in the course under review and in other courses. In short, percentiles need to be interpreted with care and in context and, in the process, they can best be regarded as identifying about five levels, not 99 levels, of student satisfaction. These five may be thought of as the top and bottom 10 percent, the next 20 percent (fairly good or fairly poor), and a middle 40% whose ratings are essentially average. Somewhat finer discriminations may be possible, but especially near the middle of the range, fine discriminations are suspect. A 40th, a 50th, and a 60th percentile can best be treated as identical.

Finally, attention must be given to establishing and maintaining the quality of the student rating data. This is not difficult to do, yet it is not always done. All that is required is to follow a few simple rules of good testing practice. These rules come to mind:

continued on page 2

*Promotion-Salary-Tenure

P-S-T Review, continued from page 1

pick a time when students are relatively free of distracting influences and schedule the evaluation period for that time. Schedule or arrange the period so that most, preferably all, enrolled students are present. If the evaluation forms are administered by the instructor, emotionally neutral conditions should be carefully established. It is better, however, to have a colleague or a specially selected student take charge of the administration, still being careful that conditions are emotionally neutral. This second party should gather the completed evaluations under conditions that insure students' anonymity and can then transmit the "raw" ratings to MRC for processing.

This discussion of student instructional ratings and of their use in the review of individual faculty is intended primarily as practical advice to interested parties. Many philosophical and technical questions remain untouched and unanswered. Perhaps another time.

Many opportunities for significant research are missed because they come disguised as hard work.
Anon

CMS - AN UPDATE

In the last issue of this memo (Volume 2, Number 1) there was a brief announcement and description of a new service to local faculty. This service, named the Course Management System (CMS), was developed initially to meet needs usually expressed by faculty responsible for multiple section, large enrollment courses.

CMS services have expanded greatly since that report, both in terms of the number of courses or faculty served and in terms of the system's capabilities. Since additional faculty may want to consider using the service, its original and present capabilities are described below.

CMS is a computer-based service designed to spare faculty the drudgery of maintaining and manipulating voluminous records relating to students' progress, to increase the accuracy of these records, and to permit rapid reporting of students' progress or status. CMS was designed to accommodate a variety of recorded information -- quiz scores, lab performance scores, attendance, etc. The system was also designed originally with capability to prepare reports to instructors which could be separated by course section, grouped across sections, and with record entries arranged alphabetically by student name, numerically by Social Security number, by points earned (e.g., on an exam), etc.

As CMS has expanded this semester to serve a varied group of about 50 courses or instructors, several features have been added. Now CMS can also assign letter grades on the basis of stated cut-off score values, produce histograms or score distributions, calculate simple statistics (means, standard deviations, N), convert absolute scores to percentages, and delete scores on the basis of stated criteria.

Faculty interested in discussing or in using this service should make early contact with Mr. John Van Horn 'phone 2207.

PSI IN GENERAL PHYSICS: A TWO COURSE SEQUENCE

Beginning in 1973, the first of what is now two self-paced courses in general physics (220T and 221T) has been available as an option to Purdue students. The two courses use a version of the Keller Plan, or PSI, and serve as an alternative to the traditional and still popular lecture-recitation-laboratory course format. As with other PSI courses, the most distinctive features of 220T and 221T are clearly stated behavioral objectives, emphasis on mastery learning, repeatable comprehensive testing of students, absolute grading standards which are explained at the outset, and (within limits) self-pacing.

There are no regularly scheduled class meetings. Instead, students come to a special study center at times which they choose. There, they receive printed study guides (12 for each semester's course) which contain the stated objectives, a large number of exercises and problems, and detailed problem solutions. Students progress through the units by demonstrating mastery, defined as a score of 80% on the unit test; tests are administered within the study center, on request. While these tests may be repeated, students' scores count toward their final grade. Nevertheless, when a test for a given unit is repeated, the highest obtained score is the one maintained in the student's record. Thus, even though the student is only required to meet an 80% criterion, most are not satisfied with marginal scores and will repeat the unit test, usually achieving higher scores. They must, however, complete each test by a stated deadline, otherwise 10 points are subtracted from the score received. Each semester, three comprehensive examinations are also given. These are scheduled at fixed times, but may be repeated once, with the higher score counting toward the final grade. About three-fourths of the students make use of this option.

At the end of each semester, students in 220T and 221T take an examination identical to that taken by counterparts in traditionally taught course sections. Over 4 semesters, final exam scores of the 220T (PSI) and traditional 220 students have been compared, after statistically adjusting for the (small) inter-group differences in SAT-M scores and high school rank. Each of these quasi-experiments has shown consistently higher scores for students in 220T. The mean differences between group scores correspond to about one letter grade. Also, a review of student attrition data (i.e., those who transfer, drop the course, or who fail to complete the requirements) does not suggest that the higher mean scores in 220T are explainable in those terms.

Persons interested in further information may contact Prof. R. W. Stanley, Department of Physics.

Now today even Lucifer will admit, if you get him in a corner, that truth is the mightiest warrior of them all. The only trouble is truancy. Sometimes he can't be found for centuries. Then he will bob up unexpectedly, break a few heads, and skip away.
Heywood Broun

RESULTS OF THE "STRAW POLL"

About one year ago, an issue of this memo (Volume 1, Number 3) included a "Straw Poll" concerned with CAFETERIA evaluation services, how these were being utilized by faculty, and how they might be improved. A total of 267 faculty responded to this informal poll and their responses can now be summarized and some comments provided. The principal work on these results was done by W.M. Goodwin III, during a brief post-doctoral appointment in the Measurement and Research Center. Dr. Goodwin is now director of MRC's counterpart unit at the University of Miami.

Of faculty who responded to the poll, 26% described themselves as full professors, 33% as associate professors, 28% as assistant professors, 7% as instructors and 5% as teaching assistants. Eighty-one percent reported that they had used CAFETERIA to evaluate their teaching, although usage is not distributed proportionately among the academic ranks. Assistant professors constitute the largest single group of CAFETERIA users, evaluating more classes more often, lending some support to the speculation that promotion is an important factor in the decision to evaluate instruction. Moreover, 58% of the respondents favored a University policy which requires all faculty to be evaluated by students, at regular intervals, and about half of them (49.5%) favored the use of these data in promotion, salary and tenure decisions; 3% were undecided and 38% opposed this practice. No significant differences were observed here among instructional ranks and thus, for these respondents, support and/or opposition can be characterized as "broad based."

Virtually all respondents wrote comments, but space does not permit a full review of their variety and scope. Instead, several have been selected for inclusion here because they address central concerns and provide an opportunity to respond or to comment.

1. "I have heard that department heads can get CAFETERIA results for faculty in their departments."

At the time the poll was conducted, CAFETERIA results were sent to one local representative, in compliance with a vote by that department's faculty. MRC's policy, then and now, is to guarantee the confidentiality of rating data. When CAFETERIA was first implemented at Purdue, MRC was guided by a policy which permitted chairpersons to receive rating data at the request of the instructor. Since this policy requires accurate accounting of each instructor's wishes in this matter and because a single error could destroy the credibility of the more general policy which preserves confidentiality, the practice was discontinued. Two copies of each instructor's report are now generated and both are sent directly to the instructor who requested the evaluation. MRC does not and will not provide rating data to anyone but the instructor.

2. "A poor instructor can pick many questions which will not hurt him. . ."

True. Faculty can guess fairly accurately about their strengths and weaknesses and they can avoid asking about weaknesses. Nonetheless, the item selection process does not measurably affect students' responses to the five item mandatory core (see TILT, Volume 2, Number 1). The system was designed to serve self-evaluation purposes and its implementation/operation is still directed toward this goal. The faculty's active participation in evaluation remains a considerable strength of CAFETERIA, even though some faculty will choose not to utilize the system's capabilities fully.

3. "Too many people in administrative positions have trouble interpreting these ratings correctly."

This may be true. This problem is relevant to the promotion/tenure decision process and it surfaces because faculty submit CAFETERIA results for promotion/tenure consideration. MRC staff are available to consult with individuals who are building promotion cases which include the results of students' ratings. The staff is preparing for summer publication a bulletin which pertains to the administrative uses of students' ratings and which may assist faculty in establishing departmental and/or school policies.

4. "I have heard no indications that any attempt has been made to improve the original unavoidable helter-skelter of over a hundred items (in the catalog)."

In 1974, the CAFETERIA catalog was "frozen" until such time as new data would permit an informed decision concerning the (re)organization and breadth of its items. Results from statistical analyses of the catalog item data are reported in IRB 77-2, which will be released this summer.

5. "Students who get low grades tend to devalue the instructor; students who get high grades overvalue him. Any instructor can come out well in scores by giving all A's."

This was one of the most persistent doubts expressed by faculty in the straw poll. The question here is whether the instructor gives the grade, or the student earns it. Can an instructor "buy" good ratings by grading high? To find out, we correlated the proportion of students who reported an expected grade of A in each class with the instructor's median ratings on the core item, "This instructor is among the best teachers I have known." Based on 1,841 classrooms, this correlation was .32, meaning that only 10% of the variance in the ratings on this item can be accounted for by the proportion of students expecting A's. A similar study using the mean expected grade revealed a correlation of .33; in both cases, these data suggest that one might have difficulty trying to "buy" good ratings with high grades. The relationships among grades, achievement, and various ratings of the instructor and the course are surely more complex than the "quid pro quo" view suggests.

Persons interested in a more detailed report of several studies pertaining to CAFETERIA ratings may wish to obtain IRB 77-1. A limited number of copies is available from MRC, ENAD 402.

When the journey from means to end is not too long, the means themselves are enjoyed if the end is ardently desired. A boy will toil uphill with a toboggan for the sake of the few brief moments of bliss during the descent; no one has to urge him to be industrious, and however he may puff and pant he is still happy. But if instead of the immediate reward you promised him an old-age pension at seventy, his energy would very quickly flag.

Bertrand Russell

All life is an experiment. Every year if not every day we have to wager our salvation upon some prophecy based upon imperfect knowledge.

Oliver Wendell Holmes

OR

All a finite, sub-divine individual can do when acting is -- to use a term of Reichenbach -- to make a posit, or wager. The best he can do is to compromise between cues so that his posit approaches the "best bet" on the basis of all the probabilities, or past relative frequencies, of relevant interrelationships lumped together.

Egon Brunswik

GRADUATE ELECTRICAL ENGINEERING MINI-COURSES FOR ON- AND OFF-CAMPUS STUDENTS

Beginning in 1975, Purdue faculty in the areas of Communications and Solid State Materials and Devices, School of Electrical Engineering, nave developed and presented a variety of specialized mini-courses to satisfy two distinct needs. First, each mini-course treats a current topic which, while important, is not broad enough to warrant a full three-credit course. Secondly, the topics treated in the mini-courses are selected because of their currency and the importance of work in these specialized fields; as a result they are of particular interest to practicing engineers who are enrolled in continuing professional education.

The mini-courses are offered under two umbrella listings and titles, EE640, Topics in Communications Engineering, and EE650, Topics in Solid-State Materials and Devices. Either of these courses may be repeated by a student, since they represent varying subject matter. In EE640, for example, the mini-course might be titled "Rate Distortion Theory", "Error Correction Codes", "Introduction to Radar", "Optical Communications", "Pulse Code and Delta Modulation", etc. and examples from EE650 include "Magnetic Memory and Recording", "Silicon Planar Technology", "Image Scanning Devices", "Preparation and Characterization of Thin Film Circuits", etc. Typically, six of these one-credit, five week courses are offered each semester, three each under the EE640 and EE650 designations.

They have been successfully offered to both the local and traditional graduate students and to off-campus students for whom these serve as continuing education. This latter student group receives the courses via Indiana's IHETS television network and their participation in the mini-courses has typically been greater than in the more traditional, 3-credit courses. For example, enrollment in the traditional courses usually average about 20 students whereas the average mini-course enrollments are in the range of 30 to 50 students.

For additional information, contact Prof. Robert M. Anderson, Jr., School of Electrical Engineering, 216 Potter Engineering Center.

The academic labor market is an exchange where universities speculate in future prestige values, based on yet undone research.

Bruce Stewart

USING FICTION AND DRAMA TO TEACH ABOUT FAMILY RELATIONSHIPS

As an outgrowth of graduate and undergraduate courses in family life education, and with a grant from the University's Institute for Family and Consumer Studies, a 27-minute video tape, "Family Images in Fiction and Drama" has been produced. The tape has the twin purpose of encouraging teachers and student teachers to use creative literature in their family relations classrooms, and of providing students with vivid illustrations of principles of family life found in fiction and drama.

"Family Images in Fiction and Drama" employs University drama personnel to portray sections and scenes from such works as Joseph Heller's Something Happened, Edward Albee's Who's Afraid of Virginia Woolf?, Robert Anderson's Double Solitaire, Peter DeVries' Through the Field of Clover, Ingmar Bergman's Scenes from a Marriage, and Dan Wakefield's Starting Over. A narrator indicates the relationship of these scenes to current research on family topics.

The value of literary sources for the family relationship class is summarized in the film with the statement by social psychologist Ronald Fernandez that literature "forces us to come back to man", and the point is emphasized that such sources, in tandem with good research materials, should increase the learning and remembering of principles of family life.

"Family Images in Fiction and Drama" was produced through the cooperation of Purdue University's Department of Child Development and Family Studies and the Telecommunication Center. Information about the Department's effort to incorporate literary sources in its education about family relationships, and about the video tape, can be obtained from Professor Richard K. Kerckhoff, Department of Child Development and Family Studies.

At school you are not engaged so much in acquiring knowledge as in making mental efforts under criticism. A certain amount of knowledge you can indeed, with average facilities, acquire, so as to retain, nor need you regret the hours spent on much that is forgotten. for the shadow of lost knowledge at least protects you from many illusions. But you go to a great school not so much for knowledge as for arts and habits; for the habit of attention, for the art of expression, for the art of entering quickly into another person's thoughts, for the habit of submitting to censure and refutation, for the art of indicating assent or dissent in graduated terms, for the habit of regarding minute points of accuracy, for the art of working out what is possible in a given time, for taste, for discrimination, for mental courage and mental soberness.

William J. Cory

TRENDS IN LEARNING AND TEACHING

A memo series for Purdue's teaching faculty, issued once each semester by the Office of Instructional Services, Measurement and Research Center, ENAD 402

Appendix D

Item Cluster Names, Together with The Contents of Their Clustered Items

Instructor Fairness

My final grade will accurately reflect my overall performance.
Grades are assigned fairly and impartially.
My instructor deals fairly and impartially with me.
This course provides an opportunity to learn from other students.
Exams are fair.
Exams stress important points of the lectures/text.

Instructor Helpfulness

My instructor returns papers quickly enough to benefit me.
I feel free to challenge my instructor's ideas in class.
My instructor is actively helpful when students have problems.
My instructor recognizes when some students fail to comprehend.

Instructor Preparedness

My instructor speaks audibly and clearly.
My instructor holds the attention of the class.
My instructor readily maintains rapport with this class.
This instructor encourages divergent thinking.
My instructor is able to simplify difficult materials.
My instructor is careful and precise when answering questions.
My instructor displays a clear understanding of course topics.
My instructor seems well-prepared for class.

Global Instructor Effectiveness

I understand easily what my instructor is saying.
My instructor makes good use of examples and illustrations.
My instructor recognizes when some students fail to comprehend.
My instructor explains difficult material clearly.
My instructor makes learning easy and interesting.
I like the way the instructor conducts this course.
My instructor has an effective style of presentation.
Overall, this instructor is among the best teachers I have known.
Teaching methods used in this course are well chosen.
My instructor develops classroom discussion skillfully.
I would enjoy taking another course from this instructor.
My instructor stimulates interest in the course.
The climate of this class is conducive to learning.
My instructor motivates me to do my best work.
My instructor makes me feel involved with this course.
My instructor displays enthusiasm when teaching.

Global Course Challenge

My instructor emphasizes relationships between and among topics.
My instructor emphasizes conceptual understanding of material.
My instructor helps me apply theory to solve problems.
This course builds understanding of concepts and principles.
My instructor has stimulated my thinking.
This course has effectively challenged me to think.
Overall, this course is among the best I have ever taken.
I highly recommend this course.
Course assignments are interesting and stimulating.

Course Applicability

This course gives me an excellent background for further study.
The practical application of subject matter is apparent.
I can apply information/skills learned in this course.
This course material is pertinent to my professional training.

Course Format

Lecture information is highly relevant to course objectives.
Course topics are dealt with in sufficient depth.
The format of this course is appropriate to course purposes.

Clarity of Course Goals

I understand what is expected of me in this course.
The stated goals of this course are consistently pursued.
This course has clearly stated objectives.

Instructional Pace

Complexity and length of course assignments are reasonable.
I was able to keep up with the work load in this course.
The amount of material covered was reasonable.
Length and difficulty of assigned readings are reasonable.

Instructional Climate

My instructor suggests specific ways I can improve.
I feel free to ask questions in class.
There is sufficient time in class for questions and discussions.
I am free to express and explain my own views in class.
When I have a question or comment I know it will be respected.
Each student is encouraged to contribute to class learning.

Appendix E

A paper presented at the 1975 convention of the American Psychological Association, Chicago, Illinois.

Faculty Evaluation*

Warren F. Seibert
Purdue University

It seems to me that my first task is probably to show that the topic of "Faculty Evaluation" is not in conflict with the general title of the symposium, a title that speaks of "Educational Psychology in Support of the College Teacher." I say this because, if we were to try some word association games using that title as the stimulus, then "Faculty Evaluation" would surely not be a frequent associate. In fact, it might occur to no one - except our esteemed chairman and a few others of us. Instead, most academics and educators will associate faculty evaluation with veiled threats, challenges to professional autonomy, and the like, not with support. In any cases, we would generally agree that faculty evaluation does not enjoy a universally good name and it usually is not considered "in support" of the teacher.

So, I will try in the time allotted to foster associative bonds between the term and the concept evaluation, and the concept of "supporting" teachers. In the process, I need to show also that educational psychologists have important roles to play in developing faculty evaluation and I'm hoping to accomplish both purposes primarily by describing some efforts and experiences I have shared with colleagues at Purdue during the last few years.

Several of us at the Measurement and Research Center there have devised an instructional evaluation system that now enjoys a fair amount of success and that I trust can teach several lessons, so let me begin with some history, with our rationale for the system, and with a description of its main features. Finally, I will suggest a few morals that I draw from the work and will try to indicate roles that educational psychologists can play and have played.

First, we call the system CAFETERIA because it gives a prominent place to a "menu" - we call it a catalog - of instructional rating items and we designed the system in response to several needs that began to emerge a few years ago. I'd like to mention those needs first.

---

*Prepared for the 1975 APA Convention, Chicago.

For one thing, as the principal testing and applied psychology support or service unit for the University, we are the place where faculty are most likely to bring their testing and educational evaluation problems, as well as other problems and needs that are somehow related. Beginning about five years ago, an increasing amount of our work consisted of helping faculty analyze thier own "home-grown" instructional rating instruments. (This was in addition to some small service for faculty who were continuing to use the old Purdue Rating Scale for Instruction.) We noticed two principal things in all this, besides the fact that the work load was getting out of hand. First, there was much variety in the kinds of things that teachers were evaluating--rarely were two instruments very similar--and we noticed also that, as measuring devices, many instruments were poor. Frequently, items were poorly worded, they left important aspects or dimensions of teaching unevaluated, they lacked discriminating power, and they gave results that were difficult if not impossible to interpret. Besides, the whole process was not at all efficient. It took too much of a teacher's time in devising instruments and too much of our time in supporting and analyzing these one-of-a-kind instruments. So, it began to occur to us that there had to be a better way.

Several other thoughts occured at about the same time--and this was almost four years ago. We thought, even though we couldn't prove it, that an important reason why so few faculty used evaluation services was that the available options were psychologically unappealing. Besides representing extra work, which of course they do, it was especially noticeable that traditional forms of evaluation offered few positive attractions to the teacher. Such instruments are typically inflexible in their contents and they are often ill suited to courses, goals and teaching styles that depart from the norm. In one sense, this problem can be too easily over-emphasized and can leave the impression that most faculty are highly innovative and non-traditional in their teaching, which is a mistaken impression, at least in our experience. But at the same time, we need to be sensitive and supportive to the minority of faculty who are non-traditional and we should be especially sensitive

to the interests of all faculty who are trying, in subtle and non-revolutionary ways, to provide better instruction. Here, it seemed to us, was an important failing of traditional approaches, so we resolved to do something helpful not only for the non-traditional subtle teacher, but for the more strategist, who we expected would appreciate evaluations that were more sensitive to the subtleties he had worked hard to produce. Besides being an educationally sound idea, as we thought it was, this also seemed good politics.

We further believed that traditional evaluation methods were denying instructors the psychic rewards of creating something for themselves or, put another way, of participating actively in a process that might have important effects in an instructor's professional life. Also, such methods implied that someone "out there" was a self-anointed expert who could pretend to know more about "my course" and "my goals" and the climate of "my classroom" than I do. And since traditional evaluation instruments are generally weak as diagnostic instruments, their results fail to tell many faculty the things they most want to know. Namely, what exactly is wrong and what could I do to improve? If these instruments are diagnostically weak, teachers may also suspect that the real purpose is to provide general and more or less final evaluations of goodness. Related to this, faculty tend to think, sometimes with a basis in fact, that such evaluation is intended more for use by administrators or by students, rather than by the sincere and motivated teacher who will take bad news with the good and will try to change, where change seems indicated.

So, to summarize, we decided there had to be a better way when we discovered (1) the inefficiency of existing practices; (2) the interesting variety in the evaluations that faculty were seeking; (3) the traditionally low levels of faculty participation in evaluation and the near-absence of incentives to participate; and (4) the inadequacies of most evaluations as diagnostic tools.

On top of all this, we resolved to do something about problems of confidentiality. We would like to believe that this concern was a simple matter of high-minded professional ethics, but the fact is that it also is a highly practical matter. Few things can hurt an evaluation service more, it seems to us, than to have that service become a source,

or even a suspected source, of information given in privacy. So, for somewhat personal reasons we wanted the new service to succeed and for more idealistic reasons we wanted it to be helpful to colleagues, but either way, we saw no alternative to maintaining a strict confidentiality of results.

Now that I have indicated the problems as we saw them, let me indicate what our solutions are. What is the CAFETERIA system and how does it operate?

It is a flexible, computer-based system that permits each instructor to select from an item "menu" or catalog and thus "design" a unique or tailor-made diagnostic rating form for any course. The form includes a standard set of student and course demographic items, such as students' year in school, expect final grade, and so on. For local use, we adopted six items that suit our needs, but the system is designed to accommodate a great variety of others. The rating form can also include as many as 35 instructor-selected items taken from a current catalog of 200, and it finally includes five common or "core" items that are simply imposed on everyone. These are general items whose current contents focus on motivating students to learn, the instructor as an explainer, the quality of course assignments, and the overally quality of both the course and instructor.

I should add that the current item catalog is a third or fourth generation effort that has evolved from a smaller set of generally weaker items. On the whole, items in the current group provide good coverage of the domains of interest and good statistics. However, we settled on this stablized catalog not because it is ideal, but because we needed a breathing spell that was free of the complications that constant revisions can cause. Nevertheless, the system remains flexible with respect to the catalog and the core items. The catalog is designed for easy expansion to the level of 300 items, then other multiples of 300 can be added, if needed; also, the set of core items can be larger than the current set of five, or smaller, or can be supplemented by special departmental or school cores, when departments or schools adopt these, as some of ours have.

In operating the service, we send invitations to all faculty each semester and we include a set of brief instructions, an item catalog,

and an item selection form. Then, using the selection form, an instructor chooses items and indicates how many rating forms are needed. This request is then processed and the rating forms are machine-generated on a computer line printer. The instructor then receives the materials, administers the forms to the students, and returns their responses for analysis. Incidentally, although we currently get responses on response cards that are separate from the questionnaire, an alternative version of the system uses a machine-printed, machine-readable questionnaire that serves as a turn-around document. This is a more elegant kind of system and in some respects preferable, although not uniformly better.

In any event, the students' responses are analyzed by computer and two copies of the confidential instructor report are prepared. This report identifies the instructor, the course, the number of students participating, and the demographic make-up of the student group. Then, for each rating item, it prints out the item statement and records the number of students who strongly agree, agree, are undecided, disagree, or strongly disagree with it; it gives the median of the distribution of responses, and it converts this median into a percentile equivalent. Both copies of the report are then mailed to the instructor.

Besides the instructor reports, broad summary reports are also prepared each semester to indicate university-wide, school-wide, or even department-wide uses of the system. These reports preserve anonymity, but they show such things as the number of times each item was selected, the distribution of classroom medians, and percentile equivalents for each median level. These broad reports are helpful to us in monitoring the quality and quantity of service and they seem to find use, for example, at the school level in monitoring strengths and weaknesses of a school's educational programs.

Let me digress for just a minute to sketch the technical side of the system. As already indicated, it is computer-based and it uses three or four standard forms and materials - the item catalog, the item selection form, formatted questionnaires, and so forth. Beyond these things, CAFETERIA can be thought of as four rather efficient

FORTRAN programs that do the work of generating questionnaires, of preparing instructor reports, of generating item statistics, and of merging norm files.

The programs are designed to run in most computer environments, from large-scale systems to mini-computers and indications are that the design is successful, since several institutions are now operational.

Honesty requires that I mention two compromises we made with the main purposes and rules of CAFETERIA. First, we included the five mandatory core items, even though these are general, global items with little direct diagnostic value; however, they do provide a minimal indication of the general quality of all participating instructors and they permit any interested instructor to submit their evaluations - voluntarily - for peer review. Our rough estimate is that half the participating faculty forward their evaluation reports to departmental promotion and salary review committees.

The second compromise is the though confidential service is normally given directly to the instructor, we made provision for a special exception. If an instructor is willing to have a second person receive his or her confidential report directly, then the instructor gives the raw data - the student response cards or sheets - to the second person, who gives them to us, and we then provide the confidential report to the same (second) person. So the rule remains simple: whoever supplies raw data for processing will directly receive the confidential report. This option is exercised by perhaps ten or fifteen percent of the participants, largely in cases where teaching assistants are supervised and trained through special departmental programs.

Different people will find different morals in the story of CAFETERIA's development and will assign varying importance to the educational psychologist roles that are played. My own guess is that we have made and continue to make unique contributions of several kinds. First, as educational psychologists, we are familiar, as few others are, with the long history of student instructional rating methods and with advantages and disadvantages of alternative methods. We had similar contributions to make in the area of psychometrics and test development--and can see much developmental work that remains to be done. It seems

especially important to have recognized diagnostic needs and at the same time to have seen that the normal incentives to faculty were inadequate, but that practical ways existed to encourage participation. Perhaps most importantly of all, education psychologists have contributions to make that are based on some understanding of teaching, learning, and the instructional process. So, imperfect though this knowledge is, others in other disciplines are generally less sensitive and less knowledgeable. And if we are to contribute to practical improvements in college and university education, much surely depends on developing mechanisms for sharing and interpreting these still-limited understandings with colleagues. We think of CAFETERIA as one such mechanism and at the same time we strive to keep it in perspective by repeating one of Bill McKeachie's many wise thoughts. Bill has said--and I quote: "Teaching is still a very human and individual endeavor and its meaning is not easily captured by statistics." CAFETERIA, then, is an aid, not the answer.

So, what I have done is to describe how we began the development of a diagnostic instructional evaluation system and the kind of system that resulted. It was conceived as a service to interested, individual instructors and it draws upon several sensitivities that are clearly from the realm of educational psychology. The service has grown in three operational years to the point where we now directly serve 2500 local classrooms per year, plus at least as many from about 25 other Indiana campuses. In addition, about 15 non-Indiana institutions have adopted the system to provide service to their faculties; these range from major research universities to small, teaching-oriented liberal arts colleges and community colleges, public and private, domestic and foreign.

Predictably, we think diagnostic evaluation is the right way to go. Our experience with it has been consistently reassuring and we see much hope for its future development. S3 A

Appendix F

A paper presented at the 1977 convention of the American Educational Research Association, New York City.

Strengths and Vulnerabilities Of the CAFETERIA Model[1]

J. O. Derry
Measurement and Research Center
Purdue University

In the past, educational researchers have warned against efforts to combine into a single evaluation system programs to improve instruction and programs to support administrative decision-making. These persons cite two reasons: First, improvement requires diagnosis, the identification of strengths and weaknesses within the instructional process, whereas administrative decision-making mandates judgments about instructional worth or value. The distinction between purposes follows the broader considerations of formative and summative evaluation, which dictate differences in the choice of items comprising the rating instrument and in the interpretation of results. Diagnosis or formative evaluation requires the flexible application of measures selected for their relevance to each instructional situation; administrative decision-makers typically want comparative data and thus they need global items broad enough to be universally applicable but powerful enough to be consistently discriminating.

The second reason cited is the lack of functional compatibility between programs to improve and to judge. Certain political considerations are involved, not only for the faculty who may collect data for

---

[1] Paper delivered at AERA National Convention, New York, April 1977, for the symposium, "Teaching, Academic Programs, and Ratings: To Improve or To Judge?"

This paper describes Purdue's CAFETERIA system for course and instructor evaluation. CAFETERIA was developed by the Staff of the Measurement and Research Center, with primary contributions by Warren F. Seibert, Allan R. Starry, John W. Van Horn, Gary L. Wright, and the author.

Supported in part by grants from Lilly Endowment, Inc. and the Fund for the Improvement of Post-Secondary Education.

IRB 77-1
Measurement and
Research Center
Purdue University
W. Lafayette IN

one purpose and the administration for another, but also for persons and units who deliver evaluation services to the faculty. Service units who deliver evaluation services to the faculty are likely caught between conflicting demands for confidentiality and for disclosure of information and their credibility threatened on both sides.

Nonetheless, in 1972 the staff of the Measurement and Research Center, Purdue University, created the CAFETERIA system for course and instructor evaluation. This system holds some promise for diagnostic evaluation while, at the same time, it supports administrative decision-making. The reconciliation of seemingly incompatible evaluation purposes has necessitated compromise; the adequacy of these compromises with respect to summative and formative evaluation will be discussed here. Research results which bear on important questions will be summarized.

The CAFETERIA System

Purdue's CAFETERIA system utilizes students' ratings of instruction; the unit of analysis is typically the individual classroom, although the system is not restricted to classrooms. Perhaps we should conceptualize a CAFETERIA model based upon the CAFETERIA system; the latter is a trademarked equivalent of Henry Ford's Model T, an inexpensive, mass-produced evaluation system which can be installed at virtually any institution. The former is a concept and a set of procedures by which the face validity of rating instruments can be substantially improved by the simple device of making each rating form more applicable to instructional goals, settings and strategies.

CAFETERIA evolved along practical considerations, with the user fully in mind. As an initial step, we collected from among our faculty as many different ratings forms and items as we could find. These we edited to conform to a single response format; because a Likert-type scale is flexible and because it was one most often used by the "home-made" rating forms we located, it was adopted. The edited items, plus numerous others written by MRC's staff, were organized into content

categories, assigned numbers serially, and distributed to the faculty in the form of a catalog. Faculty were invited to participate in what was then called "our experiment". After three semesters' experience, we designed the present CAFETERIA system.[2]

With CAFETERIA, instructors are able to tailor evaluation forms according to what each perceives as relevant. Each instructor persuses the 200-item catalog, selects those items of interest and marks these selections on a machine-readable paper form. These request forms are input to the computer, which assembles and prints the rating instrument in quantity. This "tailoring" process facilitates diagnostic evaluation, a strength of the system which presumably enhances the utility of students' ratings. Rating forms are distributed to students; when filled out they are returned for scoring. An instrutor report is generated, listing items, student response distributions, medians and percentile equivalents.

To assist with interpretation, each item is normed, separately, based upon data from its past uses. Since these norms are highly susceptible to sampling biases and because the catalog items tend to be narrow, a mandatory core consisting of five global items is added automatically to each rating form. These core items guarantee that each of three dimensions--student motivation, interest in assignments and clarity of the instructor's presentations--plus two overall assessments, of the course and of the instructor, are represented. As with catalog items, each core item is normed separately, based upon historical data across the institution, without regard to academic department, discipline, instructor rank, etc. However, CAFETERIA can generate and

[2]CAFETERIA consists of four FORTRAN computer programs which direct printing of rating forms, score and report item data, generate overall summaries for groups of instructors and courses, and builds and updates norm libraries. The programs are written without machine specific subroutines.

retrieve separate norms for any relevant demographics or demographic crossbreaks. The utility of such "discriminated" norms remains an unresolved issue among our staff.[3]

Although CAFETERIA combines, it does not confuse evaluative purposes, that is, it produces information used for diagnosing instructional strengths and weaknesses (through the use of its catalog) and it yields information that is useful in administrative decision-making (with its mandatory core). We believe that this solution to the conflicting demands of summative and formative evaluation allows purposes to complement one another. The core provides instructors with an evaluative "base line" against which they can interpret the diagnostic profile and, secondly, instructors can submit data from the core items in support of promotion and tenure cases. Parenthetically, the climate which surrounds the use of instructional ratings at an institution will affect CAFETERIA's balance between formative and summative evaluation. In designing CAFETERIA, we allotted up to ten spaces, out of forty-five which are available on our student response form, for the institution's core. However, to use the remaining thirty-five spaces to maximum advantage, faculty must be free to select items which tap potential weaknesses as well as those likely to reveal strengths. For this reason, we believe that CAFETERIA ratings should remain largely confidential, whether or not the collection of such ratings is made mandatory. A fundamental distinction can be drawn between the routine review of all rating data by administrators and/or faculty committees and the submission of relevant core data by the instructor in support of his or her promotion case.

---

[3] Discriminated norms insure that like things (people and courses) are compared. Such norms also condemn one-half of all "like things" to the tailing, negative end of the distribution.

Using the Item Selection Process To Bias Evaluation Results

Early in this paper, it was suggested that CAFETERIA represents one approach to improving the face validity of students' ratings. Intuitively, the utility of rating for diagnostic purposes is dependent upon the range and variety of cataloged items and upon the care each instructor exercises when selecting from among these. We believe that faculty are good judges of item applicability and that most will select purposefully. Although the Purdue faculty, as well as the faculty of other institutions, have agreed, several have suggested that the item selection process, supposedly a strength of the system, permits instructors to bias evaluation results by avoiding items on which unfavorable student responses are anticipated. Such biasing would be a serious system weakness if it influenced student responses to the core items. These core items always appear as the last five items on the rating form.

If we believe that faculty can judge item applicability, we must also believe that faculty can distort the diagnostic profile by omitting potentially unfavorable items. If such biasing creates a response set which extends to the core items, the use of these data to compare instructors and/or courses should be seriously challenged. In order to satisfy the requirements of formative evaluation, we instituted the item selection process, a solution which may, critics charge, obviate CAFETERIA's summative capabilities. By way of reply, we have cited the rectangular-appearing response distributions (of classroom medians) which characterize the five core items and compared these with the more skewed distributions typical of most catalog items. If biasing does occur, therefore, its effect on core items is likely minimal.

Nonetheless, Ranney (1977) asked colleagues at the University of Tennessee to create for each of their classes two rating form versions. First, each colleague selected only those items on which unfavorable ratings were expected, then each selected a similar number of items on which favorable responses were anticipated. To each "favorable" and

"unfavorable" rating form, the five item core was added; thus constituted, the rating instruments were administered to each of the classes in random sequence. Ranney's results indicate that these instructors did create positive and negative biases in their diagnostic profiles; however, these biases did not measurably influence ratings on the five core items.

From these results, we conclude tentatively that CAFETERIA is vulnerable to rigging of the type demonstrated by Ranney. Favorable or unfavorable biases can result from the item selection process and these biases have potential for damaging the quality of formative evaluation. However, summative evaluation appears not to be affected and thus, the "goodness" of CAFETERIA's compromise between summative and formative requirements is not jeopardized thus far. Here faculty's active participation in evaluation remains a considerable strength, even though some faculty will choose not to utilize the system to its fullest capabilities.

Summative and Formative Considerations When Using Catalog and Core Item Norms

Norms pose especially vexing problems. Generally speaking, in order to build item norms, items must produce variability in responses; the item, "this teacher manifests characteristics of persons not deceased," wastes time, space and money. It is frivolous; it produces no sensible variance. Variability in students' ratings can result from differences observed in the referrent, differences among student raters and differences attributable to extraneous sources. For diagnostic evaluation, each source of variability--referrent, rater, and extraneous source--is important when interpreting results, whereas for summative evaluation, we typically assign all but referrent variability to an error component (Marsh, et. al., 1975; Doyle and Whitely, 1974; Rodin, 1973).

On the other hand, variability can be obtained at validity's expense. Under ideal sampling conditions, many rating items produce a wide response range and lend themselves to norm-referenced interpreta-

tions. Whether any single item produces a response distribution appropriate to this purpose depends, to some degree, upon its wording. Items can be written and rewritten until the desired response distribution is obtained. However, there exists instances where gradations among referrents are difficult to measure or when such gradations are irrelevant. For example, death is a highly intransigent condition, although several stages and definitions are possible medically. Unless we are intimately involved, we generally accept a single condition as a telling criterion. In writing catalog items and when analyzing their response distributions, we do well to remember that not all instructional activities and teacher attributes will produce distributions with many gradations. For some items, gradations are not perceived or are not important to student raters, although the overall distribution of instructor medians may give the illusion of such gradations.

Figure 1, for example, shows the distribution of median responses for 1,770 instructors who selected the item, "My instructor displays a clear understanding of course topics." This item is the most highly skewed of any in the catalog. If the intent which justifies this item is to obtain a rank ordering of university professors according to their knowledge of their respective subjects, students are not the best source of information. If, however, we want to identify aberrant cases, those which deviate from expectation and perhaps contribute to instructional problems, the item has warrant. Its reliability estimates are moderate to high: convergence (that is, agreement on median rating for an instructor by two classes across, in this instance, fifty paired sections of a multisectioned course) was .76; split-half for median stability was .84, corrected; its commonality estimate was .86; its row maximum correlation within its content category was .80. The lowest reliability estimate ($r = .70$) was obtained by correlating median scores at semester's middle and end for approximately 110 sections of introductory "science".

FIGURE 1
DISTRIBUTION OF MEDIAN RESPONSES
FOR SAMPLE CATALOG ITEM

CATALOG ITEM 2 WAS USED 1770 TIMES.
MY INSTRUCTOR DISPLAYS A CLEAR UNDERSTANDING OF COURSE TOPICS.

	SCORE	FREQ	PCTILE	----I----I----I----I----I----I
	1.0	0	1	
	1.1	0	1	
	1.2	0	1	
	1.3	0	1	
	1.4	0	1	
	1.5	0	1	
	1.6	0	1	
	1.7	0	1	
	1.8	0	1	
	1.9	0	1	
	2.0	0	1	
	2.1	1	1	*
	2.2	0	1	
	2.3	3	1	*
	2.4	1	1	*
	2.5	1	1	*
	2.6	0	1	
	2.7	0	1	
	2.8	3	1	*
	2.9	1	1	*
	3.0	3	1	*
	3.1	1	1	*
	3.2	2	1	*
	3.3	1	1	*
	3.4	2	1	*
	3.5	16	2	*
	3.6	5	2	*
	3.7	13	3	*
	3.8	32	5	**
	3.9	47	7	***
	4.0	71	11	****
lower 25%	4.1	88	16	*****
	4.2	82	21	*****
	4.3	136	29	********
	4.4	96	34	*****
mid 50%	4.5	143	42	********
	4.6	133	50	********
	4.7	171	59	**********
	4.8	308	77	********************
	4.9	208	93	**************
	5.0	130	99	*******

Score indicates median values between 1.0 and 5.0, at .1 intervals;
Freq indicates the frequency with which each median score was obtained for 1,770 classroom uses;
Pctile indicates the percentile equivalent for each median score;
The (*) shows the per cent of the total falling into each median score category.

Nonetheless, the item shown in Figure 1 does not lend itself to a norm-referenced interpretation; perhaps it is properly considered a criterion having some undetermined cutting score. We see that percentile norms for this item are highly sensitive to small changes in median scores, so, for example, an instructor who receives a median of 4.7 from ratings in one class and 4.9, from ratings in another, will be ranked at the 59th and the 93rd percentiles, respectively. To illustrate further: suppose these ratings were based upon responses from two classes of twenty students each and every student agrees that this instructor displays a clear understanding of course topics (i.e., marks response positions 4 or 5).

	(5)	(4)	(3)	(2)	(1)		
Class #1	16	4	0	0	0	Mdn=4.9	93%tile
Class #2	12	8	0	0	0	Mdn=4.7	59%tile

The question here is not whether the responses to this item in each class are reliable, but whether the use of percentile rankings, in this instance, facilitates interpretation. Both classes describe the instructor similarly and in neither case should we assume that instructor "knowledge" is anything but satisfactory. We must conclude that at the upper end of this distribution the item does not produce meaningful discriminations, either for summative or for formative purposes.

On the other hand, CAFETERIA's core items have more rectangular-appearing distributions. In Figure 2, medians for 4,040 classrooms are plotted for core item #4, This course is among the best I have taken. We intended that this item serve summative purposes, and we selected it from among others having similar wording on the basis of its response distribution. Norms for this and the other four course items are not especially sensitive to minor fluctuations in median scores. Using the example above:

	(5)	(4)	(3)	(2)	(1)		
Class #1	16	4	0	0	0	Mdn=4.9	99%tile
Class #2	12	8	0	0	0	Mdn=4.7	96%tile

FIGURE 2
DISTRIBUTION OF MEDIAN RESPONSES
FOR SAMPLE CORE ITEM

CORE ITEM 4 WAS USED 4040 TIMES
OVERALL, I CONSIDER THIS COURSE AMONG THE BEST I HAVE TAKEN.

	SCORE	FREQ	PCTILE	----I----I----I----I----I----I
	1.0	2	1	*
	1.1	1	1	*
	1.2	5	1	*
	1.3	11	1	*
	1.4	13	1	*
	1.5	21	1	*
	1.6	7	1	*
	1.7	16	2	*
	1.8	39	3	*
	1.9	39	4	*
	2.0	63	5	**
	2.1	73	7	**
	2.2	71	9	**
lower 25%	2.3	119	12	***
	2.4	87	14	**
	2.5	125	17	***
	2.6	66	19	**
	2.7	124	22	***
	2.8	88	24	**
	2.9	104	27	***
	3.0	197	31	*****
	3.1	117	34	***
	3.2	189	39	*****
middle 50%	3.3	121	42	***
	3.4	117	45	***
	3.5	228	51	******
	3.6	165	55	****
	3.7	148	58	****
	3.8	293	66	*******
	3.9	215	71	*****
	4.0	265	77	*******
	4.1	149	81	****
	4.2	128	84	***
upper 25%	4.3	153	88	****
	4.4	68	90	**
	4.5	104	92	***
	4.6	77	94	**
	4.7	83	96	**
	4.8	83	98	**
	4.9	41	99	*
	5.0	25	99	*

Score indicates median values between 1.0 and 5.0 at .1 intervals;
Freq indicates the frequency with which each median score was obtained for 4,040 classroom uses.
Pctile indicates the percentile equivalent for each median score.
The (*) shows the per cent of the total falling into each median score category.

Moreover, normative gradations behave similarly throughout the response distribution, and thus disparities among norm rankings are attributed to substantial differences in median ratings. For example:

	(5)	(4)	(3)	(2)	(1)		
Class #1	12	8	0	0	0	Mdn=4.7	96%tile
Class #2	0	12	8	0	0	Mdn=3.7	58%tile
Class #3	0	0	12	8	0	Mdn=2.7	22%tile
Class #4	0	0	0	12	8	Mdn=1.7	2%tile

Thus, for the core items, percentile rankings can aid interpretation, both for summative and formative purposes. Promotion committees can assume that differences in ranks represent fairly wide differences in students' expressed satisfaction with courses. Individual instructors can compare their course ratings across time and with the ratings of other courses to estimate the probable impact of course improvement efforts, a topic which is presented in more detail later in this paper.

Convergence and Anchor-Point Hypotheses

For most catalog and all core items, the use of norms complements the information reported to each instructor by providing a guide for estimating strengths and weaknesses relative to some reference group. We have seen that the norms on the summative core are not especially sensitive to minor differences among raw score medians, so we might now ask whether major differences can be interpreted. At this point, two lines of reasoning can be pursued. First, numerous researchers have attempted to show that students' satisfaction with the course or with the instructor is some how related to students' achievement (Marsh, et. al., 1975; Frey, 1973; Rodin and Rodin, 1972; Remmers, 1949). Certainly, a finding that students' ratings covary with students' achievement would improve the interpretability of the former. Unfortunately, the evidence is not very conclusive, although typically such studies have produced low, positive correlations (.2 to .4). Or we can take the approach that achievement is but a component of a multifaceted satisfaction "variable". We need not assume that learning is unidimensional

nor that it has the same valence for all students. We might, as a matter of preference, accept the view that teachers perform many tasks in the classroom including those of lecturing, assigning, motivating and testing. Moreover, teachers have personalities and teaching styles: some are personable, others aloof; some are direct and others non-direct. Finally, students have tasks to perform, including those assigned by the instructor and those which represent the students' individual goals.

For the sake of argument, assume that students' ratings have two primary components: One summative (overall satisfaction) and one formative (a profile of instructional strengths and weaknesses). Assume also that students are, in the main, task oriented, so that ratings reflect students' satisfaction with the balance between imposed and idiosyncratic goals, between difficulty and pace of the assignments, between severity of performance standards and helpfulness of the instructor, and between expected achievement and actual achievement as evaluated by the teacher. In other words, assume that students' ratings can be every bit as complex as the activities they are evaluating.

Ranney's (1977) study provides some evidence that instructors have an inkling of their relative strengths and weaknesses before evaluations in a particular class are collected. How they came to this knowledge is not as important as the apparent concurrance with it by student raters. Thus, instructor hunches and students' perceptions do converge. Of course, most of us know this already: We are not typically surprised by rating results, although we may be dismayed by the satisfaction or lack of satisfaction reported by students and as revealed by the overall "tone" of the evaluation.

For the moment, let's hold in abeyance a discussion of relative strengths and weaknesses, and examine further the concept of an overall evaluative tone. In the sections which follow, two terms will be explained: convergence and anchor-points. For our immediate purposes, convergence is the tendancy for independent sets of student raters to agree on item ratings; anchor-points are estimates of overall satisfaction, the evaluative tone or halo which permeates ratings.

Item Convergence for Summative Purposes

Although students' ratings can be complex and are properly considered so, they should be interpretable. The first criterion of such interpretability is whether or not ratings converge. To test the convergence hypothesis, let us examine an instance in which ratings should agree for independent sets of judges. Figure 3 shows data plots representing percentile rankings for fifty-five instructors teaching two sections each of a single course. Core item #5 was selected for this Figure, since the item is one used by many promotion committees, its response distribution resembles closely that of core item #4 (see Figure 2), and for the instructor, its face validity is better. The item asks students to rate the overall satisfaction with the instructor, namely: Overall, this instructor is among the best teachers I have known. Although these instructors are teaching assistants, they have full classroom antonomy; the student raters represent a cross section of the University population, but we cannot assume that individual student differences are randomly distributed across all 110 sections. Nonetheless, the estimated r between percentile rankings was .63, and like most bivariate distributions, stability is greatest at the extremes.

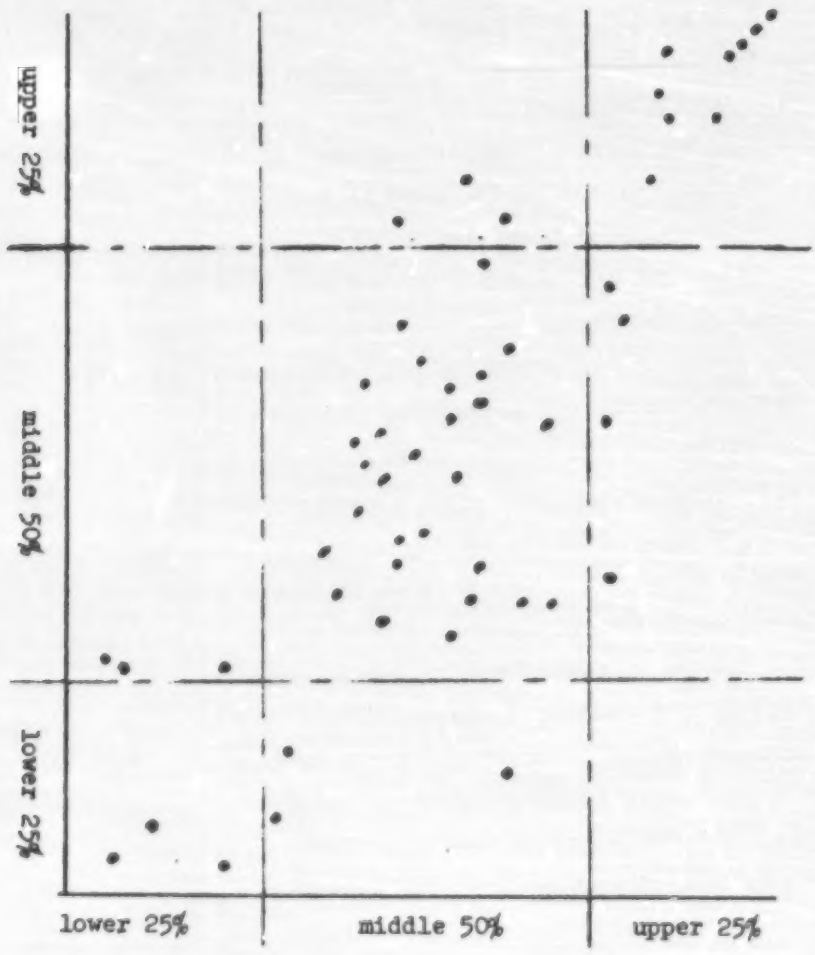
Against University norms, the highest rated of these instructors were so rated by both sections, evaluating independently. Similarly, low rated instructors fall consistently at or below the 25th percentile point on University norms. Between the 25th and the 75th percentiles, convergence is not adequate beyond the assignment of these instructors to that broad category.

For summative purposes, the use of core item #5 to rank instructors must be tempered by good judgment. First, we have no certain knowledge that the highest rated instructors are technically competent or that low rated instructors, technically incompetent. If we accept the notion that these ratings are based on students' satisfaction, then we must base our own interpretation on the possible components of satisfaction, including perceived congruence among expectations, achievement, standards, social climate, and instructor helpfulness. Second, convergence was not equally good for all instructors, although item reliability

FIGURE 3
SCATTERPLOT SHOWING CONVERGENCE
ON CORE ITEM 5

CORE ITEM 5 WAS USED 110 TIMES.

OVERALL, THIS INSTRUCTOR IS AMONG THE BEST TEACHERS I HAVE KNOWN.



Each point in the scatterplot represents ratings by two classes for each of 55 instructors teaching 110 sections of a single course. Percentile zones shown above are based upon university norms

within classes is reasonably high. For example, each core item has a conceptual relative in the catalog and the correlations between core items and their relatives range from .89 to .93 and from .91 to .94 for two semester's data. We must conclude that the use of item data (e.g., core item #5) to rank "average" teachers (those in the middle 50% on University norms) leads to tenuous conclusions. Once again, we may be imposing gradations which are not especially meaningful to student raters, and thus we may be most accurate when we conclude that student recognize the "best", the "poorest", and have difficulty placing all others.

In addition to core item #5, convergence was estimated for all other items on the rating form used for this study. Highest item convergences were found for instructor knowledge (.76), enthusiasm (.69), and effectiveness of presentation (.67); lowest convergences were for interest in the subject matter (.11) and clarity of directions for course assignments (.09). The latter items seem sensitive to individual student differences and probably hold more importance for formative than for summative purposes.

Profile Convergence for Formative Purposes

Convergence applies to formative evaluation also, and it refers to the agreement between profiles of instructional strengths and weaknesses produced by independent sets of judges. Previously, ratings were said to yield two kinds of information: overall satisfaction, perhaps estimated by core items #4 and #5, and profiles of perceived strengths and weaknesses. Of course, whether any rating on an item reports a strength or weakness is largely an interpretative matter. It is of primary concern, first, that such profiles do exist, and second, that the profiles are stable in instances where they should be stable. Here we will be concerned with raw score medians and not percentile ranks, since our purpose is formative and not summative. Because medians are the currency of CAFETERIA, and to be consistent, these and not mean scores have been reported throughout the paper, although product moment coefficients were, in fact, computed.

Profile agreement should be expected for independent sets of judges who share common rating referrents. For example, if an instructor teaches two sections of the same course, using the same syllabus, text and assignments, the two sets of ratings should produce profiles which reveal the same pattern of perceived instructional strengths and weaknesses. Even then, profiles cannot be expected to coincide exactly because some catalog items are highly sensitive to individual differences among student raters, as we saw earlier. Secondly, ratings from two sections of the same course, even though the instructors are different, should produce profiles which converge for all items for which raters have a shared referrent. Such items would necessarily pertain to the course, assignments, readings, exams (if shared), applicability of subject matter, etc. Finally, rating profiles should not converge for items dealing exclusively with instructor skills and activities for different instructors teaching the same course.

A Study to Test Profile Convergence

To test these convergence hypotheses, forty instructors who teach two sections each of a multisection course were selected. The eighty sections shared a common syllabus, text and, generally, assignments. A thirty-eight item rating form was created, consisting of nineteen items pertaining to the course and another nineteen referring to the instructor. This study was repeated during the following semester.

Based upon the average reliability of a CAFETERIA rating form ($r = .88$), the following predictions were made: if profiles for each instructor, obtained from two sections of the same course, were compared for course-referrent items only ($n = 19$), the average correlation across forty cases would approach .88, the estimated reliability of the rating instrument. Similarly, profiles for the instructor-referrent items ($n = 19$) would correlate .70, an admittedly arbitrary prediction based upon the assumption that some instructor-referrent items pertain to student/teacher relationships and hence they are likely to maximize individual student differences. These two conditions are referred to as "matched", meaning that a single instructor's ratings for two sections of the same course are compared.

For non-matched profiles, meaning a comparison between ratings for instructor A and instructor B for two sections of the same course, the predictions were these: if profiles for different instructors teaching the same course were compared, the average correlation using course-referrent items only would be .80, a second arbitrary prediction based upon the assumption that the instructor influences course profiles to some as yet unknown extent. Finally, profiles for different instructors based upon instructor-referrent items would not correlate significantly. The predictions and the study's results are presented in Table 1, and it is seen there that profile convergence followed predictions.

TABLE 1
PROFILE CONVERGENCE
FOR MULTISECTIONED COURSE

Condition	Predictions	First Study (N = 80)	Second Study (N = 74)
Profile Convergence for COURSE items (n=19)			
same instructor, same course	.88	.91**	.89**
different instructor, same course	.80	.81**	.77**
Profile Convergence for INSTRUCTOR items (n=19)			
same instructor, same course	.70	.68**	.67**
different instructor same course	.00	.37	.44

**statistically significant at .01 level

Profile convergence for "different instructor, same course" using instructor-referrent items was not statistically significant at the .05 level; all other coefficients were significant at the .01 level and beyond, as predicted. Thus, these data provide partial support for the hypothesis that profiles converge when independent groups of raters share common referrents. The stability of such profiles is essential to diagnostic evaluation, since without convergence, we are left with the suspicion that ratings mostly reflect individual student differences. However, we should remember that profiles are specific to instructors and to courses, and that much remains unknown about the interrelationships among instructor-referrents, course-referrents and raters.[4]

Anchor-points, Halo and Overall Satisfaction

As we have seen, profile convergence describes the extent to which item medians coincide for two groups of raters. A correlation is possible only when these rating profiles show similar patterns, but it does not depend upon numerical agreement at each profile point. Thus, for example, one profile can be "anchored" toward the favorable end of the response scale; the other profile, toward the unfavorable end. In other words, diagnostic strengths and weaknesses of instruction as determined from the profile are interpreted as deviations from an anchor point. Quite obviously, anchor points vary among groups of raters for the reason that ratings are highly subjective.

Although the concept of divergent anchors among judges is well known by statistians, it is not one often applied by that name to students' ratings. Typically, educational researchers refer to evaluative "halos," summarized in the extreme by the notion that liked persons do no wrong and disliked persons, no right. More formally, halo applies to the consistency of raters across supposedly independent dimensions, a rating bias which decreases item variability.

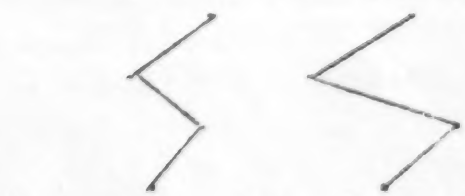
---

[4]The many contributions of past staff members are acknowledged, including especially those by Dr. Lyle Sussman, University of Pittsburg, Dr. John Weimann, Rutgers University, James Price, University of Oklahoma, and Teresa Thompson, Purdue University.

Halo is a primary component of students' ratings and throughout this paper it has been characterized as students' satisfaction with the instructional experience. Items of high specificity and low evaluative content should be less influenced by halo than are global, more evaluative items. Core items #4 and #5 were selected as barometers of rating halo, although, in fact, all five core items function in this way. Thus, the summative core anchors the diagnostic profile and in a general sense, items rated more favorably or less favorably than the core are interpreted as instructional strengths and weaknesses. When core ratings lie near the extremes of the response scale, profiles tend to be assimilated, except for a few, perhaps irrelevant, high or low points.

The anchor-point (or halo) is important to the summative interpretation of rating data. Consider the following example: In an actual case, an instructor who taught two similar courses was rated low by one class and high by another (23rd versus 80th percentiles on core item #5). Profile convergence for the two independent sets of ratings was strong ($r = .86$), indicating that both classes ranked the individual items relative to one another in about the same way. The classes apparently differed on their satisfaction with the instructional experience, although relative to these anchor-points, both classes described the instructor as hard-working, competent and enthusiastic. Similarly, both classes reported low rapport with the instructor and dissatisfaction with assignments, texts and instructional pace. Because the halo was anchored at widely divergent points on the response scale, medians for strengths in one class approached the median values of weaknesses in another. Figure 4 illustrates:

FIGURE 4
PROFILE CONVERGENCE
WHERE ANCHOR POINTS DIFFER



1.0 2.0 3.0 4.0 5.0

RESPONSE SCALE

The difference in anchor-points shown in Figure 4 and the correspondence of profiles (r = .86) in this example clearly demonstrates an important distinction between summative and formative evaluation. For the former, the instructor is seemingly erratic; for the latter, remarkably consistent, as evidenced by the stability of the profiles. In this particular instance, the two classes differed in one observable way: One class contained the usual number of students (N = 26); the other, contrary to the instructor's expectations, was larger (N = 43). We make no claim that class size influences ratings consistently, but in this instance size may have exacerbated course and instructor weaknesses or diminished the importance of instructor strengths.

Core items #4 and #5 may actually serve to estimate anchor-points for two halos, one pertaining to students' intrinsic satisfaction with the subject matter (#4) and one dealing with students' satisfaction with the instructor (#5). For the most part, these halos overlap, so that a high course rating is accompanied by a high instructor rating. In rarer instances, the halos separate in understandable if not always predictable ways: Some multi-section service courses tend to be lower rated than are the instructors; for some practicum courses, the reverse is often true.

The halo(s) in students' ratings occurs because individual raters anchor themselves at different points along the response scale, according to their satisfaction or dissatisfaction with the instructional experience. What causes students to be satisfied is not known, nor should we imagine a single cause. It is important to this discussion that we recognize the existence of such anchors in our efforts to interpret rating data.

Anchors and Profiles: Can We Change Them?

CAFETERIA was designed to meet the needs of formative and summative evaluation by bringing together two primary components of students' ratings: Expressed satisfaction with the instructor and course (as estimated by the five item core) and perceived strengths and weaknesses

of instruction (as interpreted from the rating profile). Students' satisfaction can be estimated from the "tone" of the ratings, referred to here as the anchor-point of the diagnostic profile. We have seen that anchor-point differences can occur between independent sets of raters, although such differences can accompany strong profile convergence. Whether or not anchor-points behave predictably has not been established, but it is suspected that these anchors may be more stable, especially at the extremes of the response scale, than we might ordinarily suppose.

On the other hand, students' satisfaction with instruction should not be considered a preordained thing. Hopefully, if students are dissatisfied with a course, the dissatisfaction has roots in certain observed weaknesses of instruction. Although evaluation, per se, does not improve instructors' skills, it must provide information useful to the improvement process. Of considerable importance to persons who supervise instructional improvement programs, including all those who work individually to improve their teaching, are these questions: Can rating profiles be altered? and Do such changes bring about increases in overall ratings?

Two strategies for providing answers to these questions come to mind. First, rating data from one or more courses can be obtained across time, changes can be attempted, and their impact on ratings examined. Studies of this nature are currently underway by several Purdue faculty. The other strategy involves careful monitoring of multi-section service courses and observing the effects of changes to the syllabus, text, assignments, etc. on students' ratings. Such studies are part of the on-going evaluation program at Purdue, and preliminary results can be summarized.

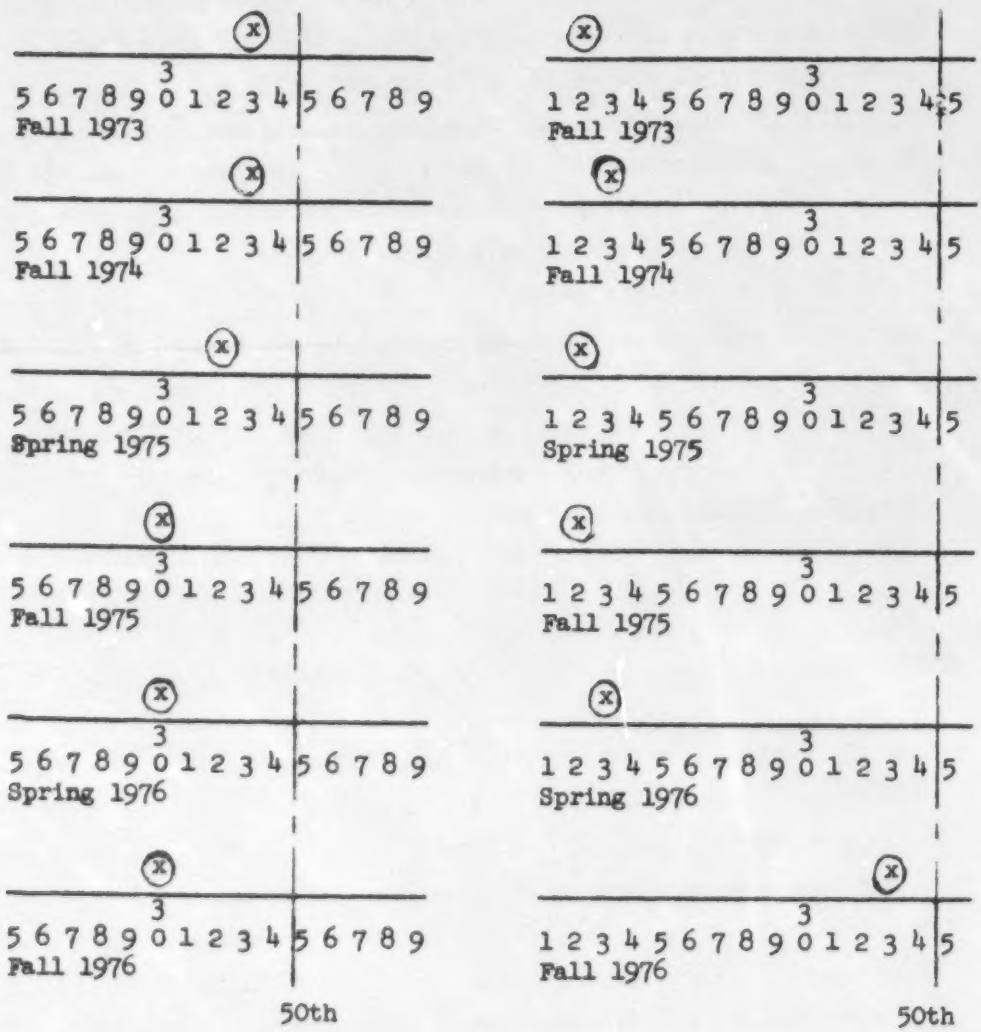
By way of background, approximately one-hundred sections of a humanities service course have been evaluated using a standardized rating instument constructed from the CAFETERIA catalog for a period spanning five semesters. These sections are organized into autonomous

classrooms and the instructors generally follow a single course syllabus, use the same textbook, and make the same assignments. From these ratings, the course's director obtains a distribution of median responses, item by item, each semester. The reporting format is a standard one provided by CAFETERIA and is shown in Figures 1 (page 8) and 2 (page 10). The median of each item's response distribution of instructor medians is the 50th percentile point; in Figure 2, for example, the overall median for core item #4 is approximately 3.5 on the response scale. The 75th percentile point lies between the scale values 3.9 and 4.0; the 25th, between 2.8 and 2.9. The course's director is able to compare medians and quartile ranges, item by item, from one semester to the next, and in this way, he monitors the continued well-being of the course.

Figure 5 shows five semester's data for core item #4, Overall, this course is among the best I have taken, and catalog item #143, I am generally pleased with the text required in this course. Other items have not been shown, but the data can be summarized: Like core item #4, item medians are remarkably stable across time. However, between Fall 1974 and Spring 1975, anchors for most items shifted downward slightly, apparently indicating a decline in students' satisfaction with the course. In an effort to improve course ratings, the director and his staff decided to correct course weaknesses as "revealed" by low points on the profile. The textbook and reading assignments were targeted for change.

Fall 1976 ratings show dramatic increases in ratings for the newly adopted textbook over previous years (see Figure 5) and minor improvements for the new reading assignments (not shown). In Figure 5, it can be seen also that course ratings remained consistent with previous years, as did the ratings on all other items, except as noted above. Thus, one part of the profile was altered, but the remaining parts, including core item ratings, stayed the same.

FIGURE 5
MEDIANS FOR A MULTISECTIONED COURSE
ACROSS FIVE SEMESTERS


Overall Course Rating
Rating of Textbook
(new text Fall 1976)
5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
Fall 1973
Fall 1974
Spring 1975
Fall 1975
Spring 1976
Fall 1976
50th

For each semester, approximately 100 sections of a multisectioned course were evaluated. The (⊗) indicates the median of the individual classroom medians. The broken line shows the 50th centile using University norms.

For the textbook item, median ratings jumped from 2.3 for Spring 1976 to 3.3 for the fall semester; interquartile ranges dropped from .9 to .6, indicating greater agreement of ratings among sections. Nonetheless, variability in median textbook ratings supports the speculation that halo effects are present and/or that instructors used the text differently in their respective courses. Some instructors are better integrated with their subject matter and show greater skill in drawing together course materials into an organized whole than do other instructors. Both explanations are valid, at least intuitively.

That ratings of the textbook improved without similar improvement in course ratings should surprise no one. In some courses the text is not a integral part clasely associated with course objectives. Certainly the course is improved whenever sources of dissatisfaction are removed or corrected. Perhaps too, the key to improvement in course or instructor ratings can be found in profile strengths, as opposed to weaknesses, or in some combination of these. As a colleague remarked, "the jury is out," and until more data are collected, the question relevant to instructional improvement and overall students' satisfaction will have to await an answer.

STRENGTHS AND VULNERABILITIES
CONCLUDING REMARKS

CAFETERIA has earned its share of criticism as well as its endorsements during the past five years. Over fifty institutions nationally use the system or a variant, but despite these uses, several fundamental questions remain unanswered. Of course, there is the continuing problem with the adequacy of the item catalog, although catalogs should never be considered complete. These should be allowed to grow as faculty find new ways to evaluate old teaching methods and as they experiment with innovative ones. Thus, CAFETERIA's strengths are its flexibility and its responsiveness to faculty's participation. On the other hand, the system is vulnerable to the climate which surrounds course and instructor evaluation and this vulnerability cannot be corrected.

Finally, CAFETERIA's flexibility extends not only to the catalog and core, but to the composition of student demographic questions as well. On this point, the system's designers, among whom the author is included, failed to appreciate that instructors want specific as well as general demographic information about students. Afterall, if ratings are measures of students' satisfaction and if each student's response contributes to the overall rating, the instructor should be able to tailor the demographic questions in ways sensible to him. The staff of the Measurement and Research Center has already collected or written numerous demographic questions for each of the institutions we serve, but we have not extended this capability to individual faculty.

The importance of demographics cannot be considered minor. Courses serve diverse student populations and among these populations, expectations, objectives and requirements can be expected to differ. If course improvement efforts are to be successful, some attention must be paid to the students served. Often, students place conflicting demands on the instructor, especially in courses which serve two or more distinct subpopulations. Consider the plight of the instructor, for example, who teaches a statistics course for one group of students who need a research tool and another group who is majoring in this subject--simultaneously, in the same classroom! These student groups may expect substantively different behaviors from the instructor; course ratings may contain different profiles and anchors for the two groups of raters. Thus, if we wish to improve students' satisfaction with instruction, then the path toward improved ratings seems to be one toward increased and not decreased sensitivity to individual student differences. As we move toward that goal, we should remember that satisfaction is probably multifaceted, so the more information we have about students the better our interpretation of rating data is likely to be.

References

Derry, J. O., Seibert, W. F., Starry, A. R., Van Horn, J. W., and Wright, G. L. The CAFETERIA System: A new approach to course and instructor evaluation. IRB 74-1, W. Lafayette, IN: MRC, ENAD 402, Purdue University.

Doyle, K. O., and Whitely, S. E. Student ratings as criteria for effective teaching. Am. Educ. Research J., 1974, 11, 259-274.

Frey, P. W. Student ratings of teaching: Validity of several rating factors. Science, 1973, 182, 83-85.

Lipshitz, R. Structural Analysis of CAFETERIA. IRB 77-2, W. Lafayette, IN: MRC, ENAD 402, Purdue University, 1977.

Marsh, H. W., Fleiner, H. and Thomas, C. S. Validity and usefulness of student evaluations of instructional quality. J. Educ. Psychol., 1975, 67, 833-839.

Ranney, G. B. An experiment to investigate possible instructor-induced item response bias in the Purdue CAFETERIA instructional evaluation system. Knoxville, TN: The University of Tennessee, 1977. Paper presented at the Seventh Annual Southeastern American Institute for Decision Sciences meeting February 1977, Birmingham, AL. Available as IRB 77-3, W. Lafayette, IN: MRC, ENAD 402, Purdue University.

Remmers, H. H., Martin, F. D., and Elliot, D. N. Are students' ratings of instructors related to their grades? Purdue University Studies in Higher Education, 1949, 66, 17-26.

Rodin, M. J. Can students evaluate good teaching? Change, 1973, 5, 66-67.

Rodin, M. and Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1164-1166.